PHYSICS

PHYSICS

FOR STUDENTS OF SCIENCE & ENGINEERING

MECHANICS AND SOUND
A. Wilmer Duff (*Editor*)

WAVE MOTION AND LIGHT
E. Percival Lewis
Revised by F. A. Jenkins

HEAT
Charles E. Mendenhall
Revised by L. R. Ingersoll

ELECTRICITY AND MAGNETISM
R. J. Stephenson

EIGHTH REVISED EDITION

630 Illustrations

THE BLAKISTON COMPANY
Philadelphia Toronto

Eighth Edition

Reprinted July, 1938
" *September*, 1938
" *November*, 1938
" *June*, 1939
" *May*, 1940
" *June*, 1941
" *January*, 1943
" *February*, 1943
" *December*, 1943
" *August*, 1946

PRINTED IN THE UNITED STATES OF AMERICA
BY HARRIS AND PARTRIDGE, INC., PHILADELPHIA, PA.

PREFACE TO THE EIGHTH EDITION

The changes in this edition of a book that has been used extensively for twenty-nine years consist of an entirely new part on Electricity and Magnetism, and a general revision of all the other parts, with a view to improving the presentation, wherever improvement seemed possible.

The part on Electricity and Magnetism in a book of this grade is the most difficult one to write, owing to the variety and complexity of the subtopics, and the numerous recent discoveries that should not be wholly passed over even in an elementary textbook. Dr. R. J. Stephenson of the University of Chicago, in writing the part on Electricity and Magnetism, has made good use of his earlier experience in teaching Physics in an engineering college. The editor is of the opinion that this new presentation, while differing occasionally from older methods, will be found stimulating and teachable.

The part on Light, originally written by the late Professor E. P. Lewis of the University of California and revised in the last edition by Professors Birge and Hall, has now been further revised by Professor F. A. Jenkins of the same university. The part on Heat, contributed by the late Professor C. E. Mendenhall of the University of Wisconsin, has been revised by Professor L. R. Ingersoll of the same university. The editor is responsible for the slight changes made in the part on Mechanics and the Properties of Matter and the more extensive changes in the part on Sound, both of which he contributed to the first edition.

In reading the page proof, primarily with a view to detecting minor slips or errors, the editor has had the valuable assistance of two friends, Professors Morton Masius and W. E. Lawton of the Worcester Polytechnic Institute. They have, however, not confined themselves to finding flaws but have offered constructive suggestions which will improve the teachability of the book.

An entirely new set of problems will be found at the end of the book; and the old problems have been extensively revised or, in some cases, replaced by ones that have been found more satisfactory.

As in the past, corrections and criticisms from those who use or examine the book will be welcomed.

THE EDITOR.

WORCESTER, MASS.

EXTRACTS FROM PREFACE TO FIRST EDITION

The preparation of a work of this grade by the collaboration of several writers is a somewhat novel undertaking, and some explanation of its genesis will not be out of place. It represents the attempt of seven experienced teachers of college physics to prepare a textbook that would be more satisfactory to all of them than any existing one. It was, of course, hoped that such a book would also prove acceptable to other teachers. It seemed to the writers that there was a need, and there would be a place, for a work prepared in this way.

One or two remarks as to the character of the book may be permitted. It will in general be found that the writers, while aiming first of all at clearness and accuracy, have preferred terseness to diffuseness. Repetition and amplification are desirable in a lecture. In a printed statement, which may be reread and weighed until mastered, they often discourage thought; and a teacher of Physics might well begin his instruction with the words of Demosthenes, "In the name of the gods I beg you to think." The writers have endeavored to present their subjects simply and directly, avoiding, on the one hand, explanations obvious to any student of fair capacity, and, on the other hand, subtle distinctions and discussions suited to more advanced courses. Some may find the material included in the book too extensive for a single course. If so, a briefer course can be arranged by omitting all of the parts in small print together with as much as those in large print as may seem desirable. There may seem to be some duplication of topics in the work of two contributors. In such cases (which are very few), it will be found that the treatment is from different points of view, appropriate to the respective subdivisions of the subject.

THE EDITOR.

WORCESTER, MASS.

CONTENTS

MECHANICS AND PROPERTIES OF MATTER

WAVE MOTION

HEAT

LIST OF TABLES

GREEK LETTERS USED AS SYMBOLS

α Alpha	η Eta	μ Mu	σ Sigma
β Beta	θ Theta	ν Nu	τ Tau
γ Gamma	κ Kappa	π Pi	ϕ Phi
δ Delta	λ Lambda	ρ Rho	ω Omega

TEXTBOOK OF PHYSICS

MECHANICS AND THE PROPERTIES OF MATTER

By A. Wilmer Duff, D. Sc.

Professor Emeritus of Physics, Worcester Polytechnic Institute.

INTRODUCTION

1. Physics as a Science.—From the evidence of our senses we infer the existence of a great variety of bodies in the physical universe around us. By the use of our senses we also learn that these bodies have various characteristics in common, such as inertia, weight, and elasticity, and these we attribute to the *matter* of which in various forms all bodies seem to consist. Matter in itself is inert; the mutual actions of bodies and the effects which they produce on our senses are due to the presence in them of something that is not matter and is called *energy*. We shall define the word energy later; the thing denoted by it is known to all, as the means which are supplied by the sun, fuels, and elevated bodies of water, and which are required for various familiar operations in nature and industry.

Physics is the Science of the Properties of Matter and Energy. This general description of Physics does not sharply distinguish it from Chemistry, and, in fact, no definite dividing line can be drawn between the two sciences, although, in a general way, it may be stated that chemistry deals chiefly with questions regarding the composition and decomposition of substances. The different branches of Engineering also treat of the properties of matter and energy, but from the point of view of their useful applications.

A science is more than a large amount of information on some subject. In very early times men must have had much valuable information regarding the physical results of various actions and

processes; but it was only when attempts were made to systematize and arrange this knowledge and to seek the relations between the different facts that the science of Physics began. The description of the phenomena of the physical world became more and more scientific, as more numerous connections between physical phenomena were discovered and described. At the present time, Physics has progressed farther in this direction than any other science, and, in seeking to give a brief account of the present state of the science of Physics, it must be our aim, not only to state the most important observed facts, but also to show the relations and interdependence of these facts.

It will be seen as we proceed that, in some parts of the subject, the relations between observed facts are better understood than in other parts. Thus in Mechanics the relations between phenomena have been so well ascertained, that we are able to start from a few simple laws regarding the motions of bodies, and from these deduce explanations of the most complicated motions. In other parts of the subject, we must be content to take, from time to time, some one principle, and trace the logical consequences of it as far as we can, and then proceed to do the same with other principles.

After classifying and studying a group of facts, the process by which we arrive at some underlying principle is called *induction*. For example, the principle of gravitation was discovered by Newton after a careful comparison of the motions of falling bodies and of the moon and the planets. Having found a general principle underlying and binding together many phenomena, we may reason forward from it and deduce other known or unknown facts, as we deduce one proposition from another in Geometry. This process is called *deduction*. In a brief account of Physics, we must necessarily use deductive more frequently than inductive methods; but, where space will permit, the effort will be made to show how, by induction, important fundamental principles have been discovered.

2. Measurement.—The first condition for success in tracing the connection between the facts in any science is that these facts shall be ascertained as accurately as possible. A qualitative statement of the size or weight of a body, to the effect that it is large or small, is of very little use. A quantitative statement of the same properties gives the ratio of the size or weight of the body to that of some accepted standard. Such a standard is called a *unit*, and the

numerical ratio of the thing measured to the unit is called the *numerical measure* or magnitude of the thing measured.

Some measurements are *direct*, that is, they are made by comparing the quantity to be measured directly with the unit of that kind, as when we find the length of a rod by placing a yard or meter scale beside it. But most measurements are *indirect*. For example, to measure the velocity of a train, we measure the distance it travels and the time required, and, by calculation, we find the number of units of velocity in the velocity of the train.

3. Observation and Experiment.—In some branches of science mere *observation*, that is, taking note of circumstances and events, is the chief or only way of obtaining knowledge. For example, the astronomer cannot modify the motions of the heavenly bodies; he must be content to observe. Observation also plays an important part in Physics, but *experiment*, which consists in modifying circumstances or events with a view to making more valuable observations, plays a more important part. Thus, if we desire to know how the earth attracts a body and whether the attraction is different at different places, we cannot make much progress if we must confine ourselves to observing bodies falling freely from various heights; but, if we modify the fall, by attaching the body to a cord and swinging it as a pendulum, we are able to make much more accurate observations and so arrive at valuable information that we could probably never gain by observing bodies that fall freely. For this reason Physics is chiefly an experimental science, that is to say, the physicist relies on carefully planned experiments to find information, and then, by methods of reasoning, and especially the condensed accurate form of reasoning called mathematics, he extracts from the results of the experiment all the information possible.

4. Hypothesis and Theory.—An event or phenomenon remains obscure or unexplained so long as its logical connection with other events or phenomena has not been traced. But it is *explained* when it is shown to be connected with other familiar phenomena and the nature of the connection is made clear. Thus, the rising of mercury in an exhausted tube was obscure and unexplained, until it was found to be different at different heights along a mountain side and to be connected with the pressure of the air on the mercury in the pool in which the tube stands. The explanation, in such a case, consists in tracing out the relation of cause and effect between

the thing explained and other things. The latter may themselves be still unexplained. Thus, the way in which air exercises pressure was not explained until comparatively recently.

A suggested explanation, so long as its correctness is still in doubt, is called an *hypothesis.* The hypothesis suggested to account for the pressure of air (or any gas) was that air consists of flying particles, which, by their bombardment of a surface, produce what we call the pressure on the surface; this suggested explanation was called the *kinetic hypothesis* of gases. The formation of a hypothesis plays a very important part in science, for it stimulates research to test its truth; and, even if this particular hypothesis is found inadequate, in testing it many new facts are usually ascertained, and the way is paved for arriving at the right explanation. The word *theory* is sometimes used in the same sense as hypothesis, but it is better to restrict it to meaning the extended discussion of an explanation or verified hypothesis. We shall use it in this sense later when speaking of the kinetic theory of gases (§227).

5. Cause and Effect.[1]—When a certain event seems inevitably to be followed by a certain other event, we are accustomed, in ordinary language, to speak of the former as the *cause* of the latter, and of the latter as the *effect* of the former. Thus the explosion of powder in a gun is spoken of as the cause of the projection of the bullet, and the latter event is described as the effect of the explosion. In speaking of the relation of two things as that of cause and effect, we do not merely mean that one has always been observed to follow the other, but we suppose that there is something invariable in the connection between them, that is, we imply our belief that nature will always act in the same way when the circumstances are the same.

6. Physical Laws.—A careful study of any phenomenon usually enables us to state, in a general way, what will happen in certain circumstances. Very ancient observation led to the conclusion that bodies, when unsupported, fall toward the earth. Such a generalization is a *physical law.* A still wider study usually leads to a more general law. Thus the study of falling bodies and of the motion of the moon and of the planets led Newton to the con-

[1] There is here no attempt to use terms in a critical philosophical sense. The use of such words cannot be avoided in an elementary work without confusing circumlocutions, and they must be used here in their ordinary sense.

clusion that each of two bodies is attracted toward the other. The aim of physical research is to obtain physical laws of increasing width and generality. Any such law is very imperfect until it can be stated in exact mathematical form, and this requires careful measurement. By measurement and calculation, Newton arrived at the law of attraction between bodies, called the Law of Universal Gravitation. **A physical law is a statement that, given a certain set of circumstances, certain events will follow.**

The proof of a physical law is sometimes *direct*, that is, the law is deduced from certain facts of observation and experiment, as one proposition is deduced from another in Geometry. Thus the law of gravitation can be derived from the known motions of the moon and the planets (§150). But in many cases the proof of a law is *indirect* and consists in showing that all results deducible from such a law are in accord with observation and experiment. This is the proof of the fundamental laws of Mechanics (§37).

7. Subdivisions of Physics.—The science of Physics may, for convenience, be divided into the following parts:

1. Mechanics.	**3. Heat.**	**5. Sound.**
2. Wave Motion.	**4. Electricity and Magnetism.**	**6. Light.**

MECHANICS

8. Mechanics is the branch of Physics that treats of the motions of bodies and the causes of changes in these motions. It is divided into two parts, one, called **Kinematics,** in which the various kinds of motion are described and studied, and the other, called **Dynamics,** in which the causes of change of motion are studied. Kinematics, or the study of motion, differs from Geometry in having to consider the element of time. Dynamics is usually divided into two parts, **Kinetics** and **Statics,** the former dealing with bodies in motion and the latter with bodies which, though acted on by causes that tend to produce motion, remain at rest, owing to the fact that these influences counteract each other. (Some authors use the term Dynamics in the sense here assigned to Kinetics.) In the following elementary treatment of Mechanics, it will not be convenient to treat the various parts of the subject quite separately; each will be taken up in turn, as convenience and simplicity may seem to dictate.

KINEMATICS

The Geometry of Displacements

9. Translation and Rotation.—Motions may be divided into two kinds. A moving body has a motion of **translation** when every straight line of particles in the body remains constant in direction. Thus a train moving on a straight track and a sled moving down a uniform incline have motions of translation. In such a case, all points in the body move in exactly the same way. Hence the motion of the body is completely described when the motion of any point in the body is stated, and we may, therefore, in describing the motion of the body, treat it as a single particle.

A body has a motion of **rotation** when all points in the body travel in circles the centers of which lie in a straight line; the line is called the *axis of rotation.* This is the motion of a grindstone, a flywheel, or a swing. Any two points in such a body are at any moment moving differently, unless they lie in a line parallel to the axis; points farther from the axis move in larger circles, and more rapidly, than those nearer to the axis.

Many forms of motion are highly complex, but they may in all cases be considered as made up of translations and rotations.

Since the motion of a body that has translation without rotation is the same as that of a point, it is convenient to begin with a study of the motion of a point.

10. Position of a Point.—The position of a point is fixed by its distances, or distances and directions, from other points, lines, or surfaces. The simplest way of stating the position of a point is by giving its distance and direction from some other point, which we may call the starting-point or *origin.*

When we confine our attention to points in a certain line, straight or curved, their positions may be assigned by giving the distance of each point from some assumed origin in that line. One direction away from the origin is taken as positive, and the opposite direction as negative. For example, the position of any station on a railway line may be fixed by its distance, positive or negative, from some other station taken as origin.

When we confine our attention to points on a surface, plane or curved, the position of each point may be assigned by its distance and direction from some origin on the surface, or, what comes to the same thing, by its distance from each of two lines that

intersect at right angles at the origin. For example, a point on the surface of the earth is described as being a certain distance east or west and a certain distance north or south from the origin.

For points not confined to any line or surface, the position of each may be assigned by its distance and direction from some assumed origin in space, or, what comes to the same thing, its distances, positive or negative, from each of three planes intersecting at right angles at the origin.

In the first case position is assigned by one number, in the second by two, and in the third by three. A point is said to have one *degree of freedom* when its motion is confined to a definite line, two degrees of freedom when it is confined to a definite surface, and three degrees of freedom when it is not restricted in any way.

The above statements of position are statements of *relative position,* that is, statements of the relation of the position of a point to that of some other point taken as origin. Absolute position, or the position of a point without any reference, stated or implied, to any other point or framework of lines, could not be described, and no definite meaning could be attached to it. In what follows the word position will always mean relative position, and, unless otherwise stated or implied, the point of reference will be some point on the surface of the earth.

11. Units of Length.—To specify a position we must use some unit of length. The unit chiefly employed in Physics is the *meter* or one of its multiples or submultiples. The meter is defined as the distance between two lines on a bar of platinum-iridium, kept at the International Bureau of Weights and Measures near Paris, when the temperature of the bar is zero centigrade. It was intended by the designers that this length should be one ten-millionth of the distance from a pole of the earth to the equator measured along a great circle. One one-hundredth of the meter is called the centimeter (0.01 m.), and this is the unit of length which we shall usually employ. Other decimal fractions of the meter are the decimeter (0.1 m.) and the millimeter (0.001 m.). For great distances the kilometer (1000 m.) is employed.

The unit of length popularly used in English-speaking countries is the *yard* or one of its well-known multiples or submultiples. The British yard is defined legally as the distance between two lines on a bronze bar kept at the office of the Exchequer in London when its

temperature is 62°F. The legal definition of the yard in the United States is $\frac{3600}{3937}$ of a meter.

12. Displacements.—A change of position is called a *displacement.* To describe it we must state its *magnitude* and its *direction.* Thus 10 miles northeast is a description of a displacement of a ship or a train. It will be noticed that we do not need to make any reference to time in describing a displacement. In fact, a displacement is a geometrical concept and does not, in itself, call for a consideration of time.

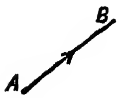


FIG. 1.—A displacement is represented by a directed line.

To help us in reasoning about displacements we draw diagrams of displacements on paper (or on a blackboard). Thus a line AB, intended to represent the displacement mentioned above, must be of such a length as to represent the magnitude of the displacement on some convenient scale, for example a centimeter to each 5 miles of the actual displacement. It must also be in the proper direction relative to lines representing other displacements. When lines are used in this way to represent displacements, AB and BA represent displacements that are equal in magnitude but opposite in direction.

For brevity we shall sometimes speak of "the displacement AB," as meaning "the displacement represented in the diagram by AB," but at other times for clearness the full expression must be used.

13. Addition of Displacements.—If a ship has sailed 10 miles east and then 10 miles north, it is $10\sqrt{2}$ miles northeast from its starting point, or we may say that the first two displacements added give the third. If one displacement of a body is represented by AB (Fig. 2) and a second by BC, the sum of the two displacements is represented by AC. By measuring the length of AC and allowing for the scale used in drawing AB and BC, we get the magnitude of the resultant displacement represented by AC. This is the *triangle method* of adding displacements.

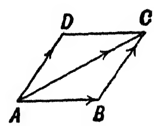


FIG. 2.

If we now draw AD equal and parallel to BC and DC equal and parallel to AB, the sum of the displacements represented by AD and DC is the displacement represented by AC, as before. Evidently, in adding displacements by a diagram of displacements, the two equal and parallel lines AD and BC are to be regarded as representing the same displacement, and similarly as regards AB and DC. In all cases *two equal and parallel lines represent the same*

displacement. This is consistent with our definition of a displacement as a *change* of position, for if two bodies move equal distances in parallel lines, the changes of position are the same, whatever their starting points may be, just as a story of given height added to a building produces the same change of height, whatever the original height may have been.

This leads us to another way of adding two displacements. If we take AB and AD as representing two displacements, their sum is the displacement represented by AC, where AC is the diagonal of the parallelogram $ABCD$ that passes through A. This is called the *parallelogram method* of adding displacements. Essentially it is the same as the triangle method, and it makes no difference which we use.

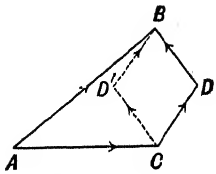


FIG. 3.—Geometrical addition of displacements.

To add three displacements AC, CD, DB (Fig. 3) we might first find the sum of the first two and then to that add the third, using the triangle or parallelogram method. A calculation of the resultant displacement would be somewhat complex. Graphically, the matter is very simple. All that we need to do is to draw a polygon of which AC, CD, DB are successive sides. The sum is the displacement AB. The result is independent of the order in which the displacements are added, as is indicated by the dotted lines in the diagram.

In the preceding we have been speaking of successive displacements. But displacements may take place simultaneously. The methods of addition are, however, the same in the two cases. For example, if a man walks 5 feet across the deck of a steamer that in the same time moves forward 10 feet, the man's final displacement is the same as if the events had followed each other.

Addition of displacements is evidently different from addition in arithmetic or algebra. The sum of two displacements each of magnitude 10 is not 20, unless they are in the same direction. The methods of adding displacements are geometrical, consisting in the construction of a triangle, parallelogram or polygon. These are the methods appropriate to the addition of quantities that have both magnitude and direction, and such quantities, as we shall see, are numerous.

If we apply the triangle method to adding two displacements AB and BA (Fig. 2), we must turn BC around until C coincides

with A. In this case the resultant becomes zero, or $AB + BA = 0$. Thus, as regards displacements, $BA = -AB$.

14. Resolution and Subtraction of Displacements.—As we may replace any number of displacements by their geometrical sum or resultant, so **we may replace a displacement by any number of displacements that, added together, give the original displacement.** This is called *resolving a displacement into components.* Thus, to resolve a displacement AC (Fig. 2) into two components in given directions, we draw from A lines in the given directions, and then complete the parallelogram $ABCD$ with AC as a diagonal; AB and AD are the components desired, since their sum is AC.

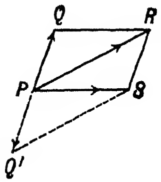


FIG. 4.—Subtraction of a displacement.

Subtraction is the opposite of addition. To subtract 4 from 10 we must find the number, 6, that, added to 4, will give 10. Similarly, to subtract a displacement, PQ, from another, PR (Fig. 4), we must find the displacement that, added to PQ, will give PR. From the triangle method of addition this is evidently QR, or, if we complete a parallelogram $PQRS$, it is PS, which is equal to QR.

Subtraction may also be performed in a slightly different way. From the triangle method it is evident that QP added to PR will give QR. Hence to subtract a displacement we may *reverse its direction and add.* This addition may be performed by a parallelogram $PQ'SR$, where $PQ' = QP$. The result is the displacement PS, as before.

15. Vector and Scalar Quantities.—Displacements belong to the class of quantities called **vector quantities,** that is, quantities that have *magnitude* and *direction.* Other vector quantities are velocities, forces, etc. In adding a number of velocities or a number of forces the methods to be used are those that we have employed for displacements.

Quantities that imply no reference to direction are called **scalar** quantities. Such are mass, volume, etc. Each of these quantities is specified by a number, without any idea of direction associated with it, and the addition or subtraction of such quantities is performed as in ordinary arithmetic or algebra.

VELOCITY

16. Velocity is rate of change of position or *rate of displacement.* Since a displacement has a definite direction, as well as a definite

magnitude, a velocity also has a definite direction and a definite magnitude, or *velocities are vector quantities.* Thus "twenty miles per hour" is not a complete statement of a velocity, since it gives only the magnitude of the velocity, and does not specify its direction; but "twenty miles per hour eastward" is a complete statement of a velocity. For clearness, such a phrase as "twenty miles per hour" may be called the statement of a *speed,* which means the mere magnitude of a velocity or a rate of change of position, without reference to the direction of the change. When the motions considered are all in the same straight line, we do not need to distinguish speed and velocity.

17. Constant Velocity.—The velocity of a point is described as *constant* or *uniform* when the displacements of the point in all equal intervals of time are equal. By equal displacements must be understood displacements equal *in both magnitude and direction.* Hence, when the velocity of a point is constant, the point moves in a straight line. The magnitude of a *constant* velocity is the magnitude of the displacement in each unit of time. Hence, if we denote the magnitude of a *constant* velocity by v and that of the displacement in time t by s,

$$s = vt.$$

Unit velocity is the velocity of a point that travels unit distance in unit time, *e.g.,* 1 cm. in 1 sec., or 1 cm. per sec. or, more briefly, 1 cm./sec. As there is no particular name for the unit, the expression cm./sec. may be used as the equivalent of a name.

18. Variable Velocity.—A point has a *variable velocity* when its displacements in equal times are not equal. The displacements in successive equal intervals of time may differ (1) in *magnitude only,* as when a point moves in a straight line with varying speed, or (2) in *direction only,* as when a point moves in a curve with constant speed, or (3) in *both magnitude and direction,* as when a point moves in a curve with varying speed. We shall begin by considering the first of these cases, that of rectilinear motion.

19. Average and Instantaneous Velocity.—In *rectilinear motion with variable velocity,* how shall we specify the magnitude of the velocity? In this case there are two ways open to us. If we divide the whole distance traversed in a certain interval of time by the length of the interval, we get the *average velocity* in that interval. If, for example, we find the whole time required by a

train to move from one station to another on a straight track, and divide this into the whole distance, we get the *average velocity* between the two stations. If the whole distance be denoted by s, the whole time by t, and the average velocity by $\bar{v}$, then $\bar{v} = s/t$. Hence

$$s = \bar{v}t.$$

The magnitude of the average velocity in an interval tells us nothing as to the way in which the motion varies during the interval. If we need to know the character of the motion more closely, we must divide the whole interval into parts and ascertain the average velocity in each. The smaller these parts, the more nearly does the average velocity in any one part represent the actual rate of motion at any moment in that part. Let us fix our attention on a certain moment at a time t after the beginning of the whole interval. If we proceeded to find the average velocity in a short interval, say Δt, following that moment, and if we took successive decreasing values for Δt, and found the average velocity in each of these decreasing values of Δt, we would find that the average velocity would rapidly approach a definite limiting value. This limiting value is the *instantaneous velocity* at the moment t. Stated more briefly, *if Δs is the displacement in a small interval of time, Δt, following the time t, the instantaneous velocity at the time t is the limiting value approached by $\Delta s/\Delta t$ as Δt approaches zero.* This may also be further abbreviated to the forms:

$$v = \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} = \frac{ds}{dt},$$

the last abbreviation being that used in the differential calculus.

When the velocity of a point is constant, the instantaneous velocity, as defined above, is the same as the velocity of the point, as defined in §17. For the values of $\Delta s/\Delta t$ at different moments in any interval t are equal. Hence, if s is the whole distance traversed in the time t, each value of $\Delta s/\Delta t$ is equal to s/t, which is the distance traversed in unit time.

When the instantaneous velocity of a point, as just defined, is variable, we may also state its magnitude in terms of an equal constant velocity. Suppose that, when the instantaneous velocity is v, the point begins to move with a constant velocity equal to v. The magnitude of this constant velocity is the distance the point

would travel in unit time. Hence we may also say that *the magnitude of the instantaneous velocity of a point is equal to the distance the point would travel in unit time, if it had an equal constant velocity.*

20. Unit of Time.—To measure or specify a velocity we must use some *unit of time.* The unit of time usually employed in Physics is the **mean solar second.** This is defined as $\frac{1}{86400}$ of the mean solar day, which is the average, throughout a year, of the time between two successive transits of the sun across the meridian at any place. It is the second of the ordinary clock or watch, when properly regulated.

21. Curvilinear Motion.—When the displacements of a point in successive equal intervals are in different directions, the point is moving in some curved path. This, for example, is the case when a ball is thrown obliquely upward or when a train is moving on a curved track. If the position of the point at a certain time t is P and at a somewhat later time, say $(t + \Delta t)$, is Q, the displacement in this time is PQ (Fig. 5). If we denote the length of the chord PQ by Δs and consider the limiting value of $\Delta s/\Delta t$, as before, we get the instantaneous velocity of the point at the time, t, at which the point is at P. As PQ is decreased the chord PQ finally approaches without limit to the tangent at P; hence the direction of the instantaneous velocity at P is along the tangent at P. While this is the proper meaning of the rate of displacement at P, we would get the same result, if we took Δs to mean the length of the arc PQ, and supposed it successively diminished by the approach of Q toward P; for the chord and the arc would, in the limit, have a ratio of unity.

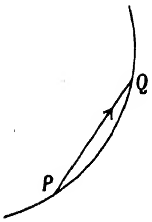


Fig. 5.

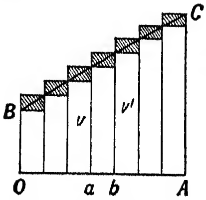


Fig. 6.—Graph of a speed.

22. Graph of Speed.—A curve drawn to represent the speed of a moving point is called a *speed curve.* Let OA (Fig. 6) be a straight line, of which the length OA stands for the length of the interval, t, in which we wish to consider the motion. Divide OA into a very large number of small equal parts. At O erect a perpendicular OB to represent the speed at the beginning of the interval t. Erect similar perpendiculars to represent the instantaneous values of the speed at the beginnings of the other parts of the interval, and draw a smooth curve, BC, through the upper ends of these perpendiculars.

Consider one of these short intervals, ab. If the speed throughout this short interval had been the same as at the beginning of the short interval, say v, the distance traversed in the short interval would have been $v \times ab$ or the unshaded rectangle above ab. If the speed throughout the short interval had been the same as that at the end of the short interval, say v', the distance would have been $v' \times ab$ or the area of the unshaded rectangle plus that of the small shaded rectangle above it. The real distance in the interval is intermediate between these two. Applying the same reasoning to all the small intervals in succession, we see that the whole distance is something between that represented by the whole unshaded area between BC and OA and that represented by the unshaded area plus the shaded area. If the number of parts into which OA is divided be doubled, there will be twice as many small shaded rectangles, but the area of each will be about one-fourth as great as before. Hence, if we suppose the number of the small intervals increased without limit, the shaded area will decrease without limit until it vanishes, and the area between the curve BC and the line OA will represent the actual distance in the time t.

Since it is merely the *magnitude* of the velocity that is represented by each ordinate, the area represents the distance *measured along the line of motion,* whether this be straight or curved.

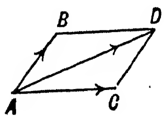


FIG. 7.

23. The Resultant of Simultaneous Velocities. A man sitting in a train has the velocity of the train, but, when he gets up and moves about, he has an additional velocity, which may or may not be in the same direction as the first velocity. Similarly, a launch floating down with the current in a river has the velocity of the current; but if it has a propeller in motion, it has another velocity, in addition to the first. *When a body has two or more simultaneous velocities, it pursues some definite path, and its velocity in the path is called the* **resultant** *of the simultaneous velocities.*

From this definition of the resultant of any number of simultaneous velocities, it follows that the magnitude and direction of the resultant velocity can be deduced from the separate velocities by the triangle, parallelogram, or polygon method of adding vectors. For if AB and AC (Fig. 7) represent two *constant* velocities, they also represent the displacements these velocities produce in unit time, and the diagonal AD therefore represents the

resultant displacement in unit time, that is, the resultant velocity.

When the component velocities are *not constant*, we can add their instantaneous values by the vector methods referred to. The proof of this statement is the same as before, except that AB, AC, and AD will now stand for the displacements that would take place in unit time, if the velocities remained constant that long.

24. Formula for Resultant.—Let v_1 and v_2 be the respective magnitudes of two component velocities of a moving point, and let these velocities be represented by OA and OB (Fig. 8). Also let v be the magnitude of the resultant velocity, which is represented by OC, where OC is the diagonal of the parallelogram of which OA and OB are sides. By a trigonometrical formula,

$$OC^2 = OA^2 + AC^2 - 2OA \cdot AC \cos OAC$$

Denote the angle AOB, which is the angle between the directions of the two components, by θ. Then the angle OAC equals $(180° - \theta)$,

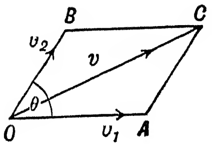


Fig. 8.

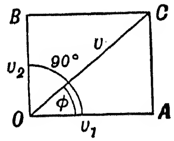


Fig. 9.

and therefore $\cos OAC = -\cos\theta$. Since OA, OB, and OC are proportional to v_1, v_2, and v, respectively,

$$v^2 = v_1^2 + v_2^2 + 2v_1v_2 \cos\theta$$

By this formula we can calculate the magnitude of v when v_1, v_2, and θ are known; a trigonometric formula for the angle AOC that its *direction* makes with OA can readily be obtained. When the components are in the same direction, $v = (v_1 + v_2)$. When they are in opposite directions, $v = (v_1 - v_2)$ or $(v_2 - v_1)$, depending on which of the components is the greater. These results can be obtained from the equation for v, but they are evident without it.

If $\theta = 90°$, that is, if the two components are at right angles, $\cos\theta = 0$ and (Fig. 9)

$$v^2 = v_1^2 + v_2^2$$

and if ϕ be used to denote the angle AOC that the resultant makes with the component of magnitude v_1,

$$\tan \phi = \frac{AC}{OA} = \frac{v_2}{v_1}$$

25. Resolution of a Velocity into Components.—Since two velocities, taken together, are equivalent to a single velocity, called their resultant, we may reverse the process, and suppose any velocity replaced by any two velocities that, when added, are equivalent to the original velocity. This is called *resolving a velocity into components.* To resolve a velocity we draw a parallelogram in which the diagonal stands for the velocity to be resolved. Now any number of parallelograms can be drawn with a given line as diagonal; but, if the directions of the sides are specified, only one solution is possible. Hence, to resolve a given velocity into components in two given directions is a definite problem.

The most important case of the above is when the directions of the components are at right angles. Thus, if the velocity is v in the direction Oc (Fig. 10), and if OC is taken to represent v and if Oa and Ob are to be the directions of the components, we draw from C perpendiculars, CA and CB, to Oa and Ob respectively. Then OB and OA are the desired components in the specified directions. If the direction Oa makes an angle θ with the direction of v, and if we denote the components in the directions Oa and Ob respectively by v_1 and v_2,

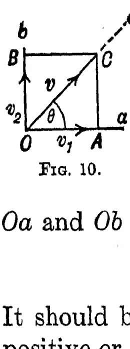


FIG. 10.

$$v_1 = v \cos \theta. \qquad v_2 = v \sin \theta.$$

It should be noted that θ stands for an angle that may be either positive or negative. We may regard θ as the angle through which a line, starting from the position Oa, must revolve about O to reach the position OC. It is customary to regard the angle as positive when the revolution is counter-clockwise, and the same applies to the right angle that Ob makes with Oa.

ACCELERATION

26. Acceleration is *rate of change of velocity.* A change of velocity has a definite direction, as well as a definite magnitude. Hence acceleration is a quantity that has both direction and magnitude, that is, acceleration is a vector quantity.

An acceleration may be either constant or variable. *The acceleration of a point is constant when the velocity of the point changes by equal amounts in equal intervals of time.* By equal changes of velocity must be understood changes of velocity that are equal in magnitude and in the same direction. When the changes of velocity in equal intervals of time are not equal, the acceleration is *variable.*

The statement that the velocity of a point is *variable* may refer to a change in the magnitude of the velocity, to a change in the direction of the velocity, or to a change in both. Hence we shall have three cases of acceleration to consider: (1) when the velocity of the point is constant in direction but variable in magnitude; (2) when the velocity is constant in magnitude but variable in direction; (3) when the velocity is variable in both magnitude and direction.

The simplest case of acceleration is when the velocity of the moving point is constant in direction and increases at a constant rate. This is illustrated by a body dropped from a height and falling in a straight line (air resistance being neglected).

The *magnitude of a constant acceleration* is the magnitude of the velocity added in each unit of time, and the direction of the acceleration is the direction of the added velocity. The *unit* of *acceleration* is that of a point the velocity of which increases by unit velocity in unit time. When the cm. is taken as unit of length and the sec. as unit of time, the unit of acceleration is such that the velocity increases by one cm. per sec. in each second, or one cm. per sec. per sec. or more briefly 1 cm/sec.^{2} As there is no name for the unit of acceleration, the expression cm./sec.^{2} may be taken as the equivalent of a name.

27. Changes of Units.—Formulae, such as cm./sec. and cm/sec.^{2}, that show the relation of some secondary or derived unit to primary or fundamental units, are useful when data are to be translated from one set of units to another. For example, if the velocity of a train is 60 miles per hour, what is it in feet per second? Suppose it is x. Then

$$x\frac{\text{ft.}}{\text{sec.}} = 60\frac{\text{mi.}}{\text{hr.}}.$$

Hence

$$x = 60\left(\frac{\text{mi.}}{\text{ft.}}\right) \times \left(\frac{\text{sec.}}{\text{hr.}}\right) = 88.$$

This result, that 60 miles per hour is equal to 88 feet per second, is worth remembering. By means of it velocities of 2, 3, 4, 5, 6, 10, 12, 15, 20, 30 miles per hour are readily translated into feet per second.

As a second example: if the acceleration of a falling body is 980 cm. per sec. per sec., what is it in meters per min. per min.? Suppose it is x. Then

$$x\frac{\text{m.}}{\text{min.}^2} = 980\frac{\text{cm.}}{\text{sec.}^2}.$$

Hence

$$x = 980\left(\frac{\text{min.}}{\text{sec.}}\right)^2 \times \left(\frac{\text{cm.}}{\text{m.}}\right) = 35{,}280.$$

28. Equations of Motion.—In considering the motion of a point along a straight line with constant acceleration, we take one direction along the line as positive and the opposite direction as negative. Let v_0 be the velocity of the point at the beginning of an interval of time of length t, and let v be its velocity at the end of the interval. The increase of velocity is $(v - v_0)$ and the increase per unit time is $(v - v_0)/t$. This is, therefore, the acceleration, which we shall denote by a. Hence

$$v = v_0 + at \tag{1}$$

This equation is simply a different way of stating that the final velocity (at the end of the time t) is equal to the initial velocity (at the beginning of t) plus the increase of velocity, and the increase of velocity is equal to the acceleration multiplied by the time.

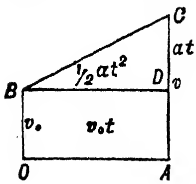


FIG. 11.

To find how far the point travels in the time t, let us consider the form of the curve of speed (§22) in the present case. The changes of speed in equal short intervals of time are equal. Hence, in Fig. 6, the differences between each ordinate and the next in order are equal, and the curve of speed is therefore a straight line, as shown in Fig. 11. Draw BD parallel to OA. The whole area $OBCA$ consists of two parts, that of the rectangle $OBDA$ and that of the triangle BCD. OB represents the initial speed, that is, the magnitude of v_0, and we shall suppose that the figure is drawn to

such a scale, that OB contains as many units of length as the speed contains units of speed, and that the same is true of AC, which represents the final speed, that is, the magnitude of v. The height of the triangle, DC, represents, in the same way, the increase of speed, which is the magnitude of at. OA represents the time t, and we shall suppose that OA contains the same number of units of length as t contains units of time. The whole area is therefore $(v_0t + \frac{1}{2}t \cdot at)$. Hence, if s is the whole distance traversed in the time t,

$$s = v_0t + \tfrac{1}{2}at^2 \tag{2}$$

This very important equation consists of two parts. The part v_0t is the distance the point would have travelled in the time t, if its speed throughout t had remained constant and equal to the initial speed v_0. The part $\frac{1}{2}at^2$ is the additional distance due to the acceleration, that is, the distance the point would have gone if it had started from rest with an acceleration a.

Between (1) and (2) we may eliminate t and so find an expression for the final velocity in terms of the initial velocity, the acceleration, and the distance.

$$\begin{aligned} v^2 &= v_0{}^2 + 2v_0at + a^2t^2 \\ &= v_0{}^2 + 2a(v_0t + \tfrac{1}{2}at^2) \\ &= v_0{}^2 + 2as \end{aligned} \tag{3}$$

Equations (1), (2), and (3) are of great importance.

Another expression for the area $OBCA$ is $\frac{1}{2}(AC + OB) \cdot OA$. Hence the distance is also given by the formula

$$s = \frac{v + v_0}{2}t \tag{4}$$

From this it follows that the average speed, which equals the total distance divided by the time (§19), is equal to one-half of the sum of the initial speed and the final speed.

Equation (2) may be readily obtained by means of the Integral Calculus. The distance travelled in a short time dt, when the velocity is v, is vdt. Hence the whole distance $s = \int_0^t vdt = \int_0^t (v_0 + at)dt = v_0t + \frac{1}{2}at^2$.

29. Galileo's Experiments.—The relations expressed by (1) and (2) of the last section were discovered by Galileo[1] by studying

[1] Galileo, "Two New Sciences," p. 173. Translated by Crew.

the motion of falling bodies, and this discovery was the beginning of Kinetics. Before that time nothing was known as to the way in which the speed of a body increases as it falls. Galileo thought the law of increase expressed by (1), namely, that the increase of speed is proportional to the time, was probably correct; but the instrumental means at his command did not enable him to test it; so he deduced (2), practically by the graphical method given in §28, and then tested it. To avoid having to deal with any great speeds, such as that of a body falling vertically, he tested the rolling of a ball down an inclined plane, assuming that both motions would follow the same law. The result confirmed his formula.

30. Acceleration of Free Fall.—We shall assume as an experimental fact, discovered by Galileo, that **at any one place all bodies falling freely would have the same acceleration, if it were not for the effect of air-friction.** The latter is very small in the case of dense solids, such as blocks of metal, falling moderate distances, and may usually be neglected. The acceleration of free fall, or the acceleration of gravity, as it is often called, is usually denoted by g. In the metric system g is about 980 cm. per sec. per sec., though slightly different at different points on the earth's surface, and in feet and seconds it is about 32.2 ft. per sec. per sec. Hence, by §28, when a body is projected vertically downward with a velocity v_0, its velocity and distance after an interval t may be found from

$$\begin{aligned} v &= v_0 + gt \\ s &= v_0t + \tfrac{1}{2}gt^2, \\ v^2 &= v_0^2 + 2gs \end{aligned}$$

When the direction of projection is upward, we may take upward as the positive direction, and g, being downward, will then be negative. In this case

$$v = v_0 - gt, \tag{1}$$
$$s = v_0t - \tfrac{1}{2}gt^2, \tag{2}$$
$$v^2 = v_0^2 - 2gs \tag{3}$$

At the highest point $v = 0$; hence from (1) we have $t = v_0/g$. Substituting this in (2), we get for the *height of ascent* $s = \frac{1}{2}v_0^2/g$. This also follows from (3), by putting $v = 0$. The time of return to the ground is got by putting $s = 0$ in (2). This gives $t = 2v_0/g$,

showing that the whole time of rise and fall equals twice the time of ascent, or that *the time of rise equals the time of fall.* It follows from (3) that the velocity of return to the starting point, that is, when s is again zero, equals the velocity of projection in magnitude, but it is in the opposite direction. It must, however, be remembered that these statements are true only for moderate velocities. At high velocities, such as those of a bullet, air-resistance greatly modifies the motion.

The value of g at any station of observation depends on the latitude of the station and also on the height of the station above sea-level. The results of very careful experiments show that, at a station in latitude λ and at an elevation of l meters above sea level,

$$g = 977.989\,(1 + .0052 \sin^2 \lambda - .0000002\, l) \text{ cm./sec.}^2$$

31. Motion of a Projectile.—When a body is thrown obliquely into the air, its motion may be considered as consisting of a horizontal part and a vertical part. The vertical part is subject to a constant acceleration g downward; while, since there is no horizontal acceleration (if we may neglect air-friction), the horizontal part of the motion is a constant velocity. If the magnitude of the velocity of projection is v and the direction of projection makes an angle θ with the horizontal, the velocity may be resolved into a component $v \cos \theta$ in a horizontal direction and a component $v \sin \theta$ in a direction vertically upward. If, then, x is the horizontal distance traversed in time t,

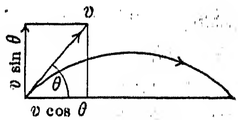


FIG. 12.—Path of a projectile.

$$x = vt \cos \theta \tag{1}$$

and if, at the time t, the vertical distance attained is y,

$$y = vt \sin \theta - \tfrac{1}{2} g t^2 \tag{2}$$

Thus the vertical motion is the same as that of a body thrown vertically upward with a velocity $v \sin \theta$. Hence (§30) at time $(v \sin \theta)/g$ the body will have just lost its vertical velocity and will therefore be moving wholly in a horizontal direction; and at that moment the height will be $(v^2 \sin^2 \theta)/2g$. At the time $(2v \sin \theta)/g$ the body will have returned to its original level and the distance horizontally from its starting-point will then be $v \cos \theta (2v \sin \theta)/g$ or $(v^2 \sin 2\theta)/g$. Now, since $\sin 2\theta$ has its maximum value, unity,

when 2θ is $90°$, that is, when θ is $45°$, it follows that the greatest horizontal range for a given velocity, v, of projection is v^2/g and is obtained by making the angle of projection $45°$.

If it be desired to find the constant relation that holds between x and y during the motion, the value of t from (1), may be substituted in (2). This gives the equation of a parabola referred to axes through the point of projection. Hence the path of the projectile is a parabola.

As in the case of §30, these results are approximately correct only in the case of the moderate velocities for which air-friction is negligible.

32. Variable Acceleration.—When the acceleration of a point is variable, we can no longer measure it by the actual increase of velocity in any time. We may, however, divide the magnitude of the increase of velocity in any time by the time and call this the magnitude of the *average acceleration* in that time, the direction of this average acceleration being the direction of the velocity added. The *instantaneous value of the acceleration* is defined much as in the case of instantaneous velocity, namely, as the value to which the average acceleration approaches as the interval is diminished without limit, or

$$a = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{dv}{dt}$$

A variable acceleration may be variable as regards magnitude or direction or both. In the following we shall consider the case of an acceleration that is constant in magnitude but variable in direction.

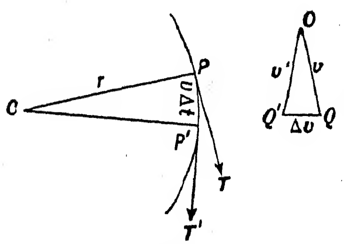


Fig. 13.

33. Acceleration of a Point Which Moves in a Circle with Constant Speed. Let P be the position of the moving point at time t and P' its position at time $t + \Delta t$. At P the point is moving in the direction of the tangent PT, and at P' in the direction of the tangent $P'T'$ (Fig. 13).

From any point, O, draw lines OQ, OQ', of equal length, to represent the velocities v and v' at P and P', respectively. QQ' will represent the velocity, Δv, added in time Δt. The triangles OQQ' and CPP' are similar. The arc PP' equals $v\Delta t$, and the chord PP'

approaches equality with the arc PP' at Δt is diminished. Hence, approximately,

$$\frac{\Delta v}{v} = \frac{v\Delta t}{r}$$

and this relation becomes more nearly exact as Δt is diminished. Hence (§32) the acceleration of the point is

$$a = \frac{v^2}{r}$$

QQ' is perpendicular to PP' and its direction approaches that of PC as a limit. Hence *the acceleration is directed toward the center.* It is therefore, *an acceleration of constant magnitude but of variable direction.*

34. Addition of Accelerations.—A moving point may have two or more accelerations simultaneously. Thus, a man at rest on the deck of a ship that is moving with an increasing speed has one acceleration, that of the ship. If he starts to move across the deck, he has a second acceleration. In any such case, the moving body travels in some path with a definite acceleration, which is called the *resultant* of the component accelerations.

The resultant of two accelerations can be found by the vector method of addition, that is, by the construction of a triangle, parallelogram or polygon, the sides of which represent the separate accelerations. For if AB and AC represent two constant accelerations possessed simultaneously by a point, they also represent velocities added in unit time. Hence the diagonal AD must represent the resultant velocity added in unit time, which is what is meant by the resultant acceleration. The same method of reasoning is applicable when the accelerations are variable; the only difference being that AB and AC and AD all represent velocities that would have been added in unit time, if the accelerations had remained constant that long.

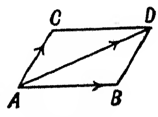


FIG. 14.

35. Resolution of an Acceleration into Components.—Since two or more accelerations may be replaced by their resultant, it follows that an acceleration may be resolved into two or more components by the ordinary methods. The case in which an acceleration is resolved into two components at right angles is especially important. The formulae for finding the components

are similar to those for resolving a velocity into two components at right angles (§25). For example, if a train is moving in a curve with varying speed, its acceleration may be resolved into a component along the tangent, which gives the rate of increase of the speed, and a component along the radius of the circle of curvature at the point, which must be equal to v^2/r (§33), where r is the radius of the circle of curvature.

DYNAMICS

FORCE AND MASS

36. In the preceding we have considered various cases of motion, without any reference to the influences that affect the motions of bodies, just as in Geometry we study lines and figures without any reference to particular bodies. We must now consider those relations between bodies on which changes of motion depend.

Isaac Newton was the first who attained clear ideas as to the relations between bodies and their motions. His treatment of the subject was founded on three fundamental principles, which he called Axioms or Laws of Motion. These axioms are so simple that they are recognized as very probably true as soon as their meaning is grasped.

37. Newton's First Law of Motion.—Every body persists in its state of rest or of uniform motion in a straight line, unless it is compelled by some force to change that state.

This law was a complete denial of what had been supposed to be true up to the time of Galileo (who died in 1642, the year in which Newton was born); for, until then, it was supposed that a body free from external influences would come to rest. No body is entirely free from external influences. Most of the stars (not the planets) are so far apart, that their motions, as measured by the astronomers, seem to remain undisturbed for hundreds of years (excepting the motions of revolution of double stars). Yet even they are probably affected somewhat by other stars. Terrestrial bodies are subject to many disturbing influences. But we can greatly diminish these, and with each diminution the (relative) velocity becomes more nearly constant. The most common hindrance to steady motion is friction. A stone given a push along a rough road is quickly stopped by friction; on a smooth

floor it will continue longer in motion; a well-polished stone started on smooth ice will continue in motion for a great distance. Such considerations make it seem probable that, if freed from external influences, a body would move with constant velocity; they do not, however, amount to a *proof* of the statement in the first law of motion. The proof of the law is that all of the innumerable deductions made from it and the other laws of motion are verified by experience.

The law implies a definition of **force. Force is the cause of acceleration.** Thus friction, the pull of a stretched spring, the attraction of the earth on a body, etc., are forces; when a body revolves in a circle, it has an acceleration toward the center and must, therefore, be acted on by some force. What exerts a force on a body is, of course, some other body. Thus the friction opposing the motion of a vehicle is due to the earth, and the pull of a towline is due to the tugboat, etc. The word "force" is therefore a

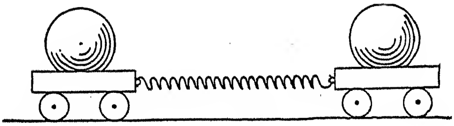


FIG. 15.

name that we give to that influence of one body on another by which the first changes or tends to change the motion of the second.

The property a body has of tending to persist in its state of motion or of rest is called *inertia*. If it were defined so as to be a measurable quantity, it would mean the same thing as mass.

38. The Mass of a Body.—Common experience shows that, when a given force is applied to a body, the magnitude of the acceleration depends on some property of the body. Thus a horizontal spring, stretched to a definite length, say one foot, will apply a definite force to the body to which it is attached. If attached to a cubic foot of lead supported without friction, it will produce a certain acceleration; but the acceleration will be different, if a cubic foot of wood be substituted for the lead. In Fig. 15 one ball is a cannon-ball and the other is a ball of wood painted black. Which is which is readily seen from their accelerations when released. The difference is not due to the difference in the weights of the bodies, since weight is a force that acts vertically and does not affect the horizontal motion of the bodies. The difference is

due to what we call the *masses* of the bodies. It might of course, be said that the difference is one of inertia, or reluctance to start or stop moving, but inertia is merely a popular qualitative term for the quantitative concept mass.

To attach a definite meaning to the word **mass** we must define what is meant by the ratio of the masses of two bodies. **The ratio of the masses of two bodies is the inverse of the ratio of the accelerations that a given force imparts to the bodies, when applied to them in succession,** that is,

$$\frac{m_1}{m_2} = \frac{a_2}{a_1}$$

when the force is the same. For example, if a body A, acted on by a certain force, receives twice the acceleration that a second body B receives, when acted on by the same force, the mass of A is half as great as the mass of B, and so for other ratios. Hence, if we adopt a certain body as a body of unit mass, the mass of any other body becomes definite.

The unit of mass chiefly employed in Physics is the *gram*, which is defined as one one-thousandth of the mass of a block of platinum, kept at Sèvres, near Paris, and known as the *Kilogram prototype*. Fractions and multiples of the gram in frequent use are named as follows:

Milligram	= .001 g.	Kilogram	= 1000 g.
Centigram	= .01 g.	Metric ton	= 1,000,000 g.
Decigram	= .1 g.		= 1000 kg.

In English-speaking countries the *pound* is, for commercial purposes, used as unit of mass. It is defined as the mass of a certain block of platinum kept at the Exchequer in London. It is worth remembering that 1 kg. = 2.20 lbs. approximately and that 1 pound = 454 gms. approximately.

We shall see later (§622) that there is good reason to believe that at immensely great velocities the mass of a body may depend appreciably on its velocity, but we shall neglect this effect for the present and consider the mass of a body as constant.

39. Ratio of Forces.—Different forces, applied to a body, give it different accelerations. For example, if a heavy body be hung from the ceiling by a cord and a horizontal cord be attached to it, a small pull will start it slowly, while a stronger pull will start it

more rapidly. Or, if a horizontal spiral spring, kept stretched to a definite length, were applied to a body supported with very little friction on a horizontal table, a definite acceleration would be produced. If this experiment were repeated with the spring stretched to a different length, a different acceleration would result. These illustrations will help to make clear the following definition of the ratio of two forces, and from this we shall be able to deduce a more accurate method of finding the ratio, either by calculation (§41) or by static experiments (§55).

The ratio of two forces is the ratio of the accelerations they can impart to a given body.

$$\frac{F_1}{F_2} = \frac{a_1}{a_2}$$

when the mass is the same. For definiteness, we shall suppose that the body referred to is one of unit mass. If we now take any force as unit force, the magnitude of any other force becomes definite. For simplicity, we usually, in Physics, take *as unit force* **that force which, acting on unit mass, gives it unit acceleration.** A force that gives unit mass two units of acceleration will then be a force of two units, and so on.

40. Momentum.—Certain properties of moving bodies depend on mass and velocity conjointly. Thus, the length of time required by a locomotive to start a train depends on both the mass of the train and the velocity to be imparted to it, and the same is true of stopping it. Hence we find it convenient to define a property depending on mass and velocity conjointly. *Momentum* is defined as *the product of mass and velocity.* Since a velocity has direction as well as magnitude, while a mass has magnitude only, the momentum of a body is a vector quantity, the direction of which is that of the velocity. (What we now call momentum Newton called *quantity of motion,* as distinguished from *rate of motion* or velocity.)

When the velocity of a body changes, its momentum also changes. Since the mass of the body is constant, any change in the momentum of a body must be due to a change of its velocity, and the change of momentum must equal the product of the mass and the change of velocity. Hence, when the momentum of a body is changing, *the rate of change of momentum equals the product of the mass and the rate of change of velocity,* that is, the product, ma, of the mass m and its acceleration a.

41. Newton's Second Law of Motion.—The rate of change of the momentum of a body is proportional to the force acting on it and is in the direction of the force.

Hence if a force F acts on a body of mass m, F is proportional to ma (§40), whatever units may be used in stating the magnitudes of F, m, and a. If we now choose a suitable set of units we may turn this proportionality into an equation and write

$$F = ma \tag{1}$$

What we mean by a suitable set of units can be explained by considering an analogous case. The volume, V, of the water that fills a tank of length x, breadth y, and depth z, is proportional to xyz, whatever units of volume and length are used. If V is expressed in gallons, while x, y, and z are in feet, V is proportional to xyz but is not equal to xyz. The condition for equality is that V must be 1, when x, y, and z are each 1. This is satisfied if V is in cubic feet and x, y, and z in feet. Similarly, in (1) F must be 1 when m is 1 and a is 1. Evidently, many different sets of units can be devised so as to satisfy this condition. Of these only two are extensively used in English-speaking countries, and they will be explained in §§42, 43.

Whatever units are used, (1) applies to the special case of a body of mass m that is falling freely and is therefore acted on only by its weight w. If its acceleration is g

$$w = mg \tag{2}$$

Now Galileo found, and Newton and Bessel proved more completely, that, at any particular place, *all bodies fall with the same acceleration* when the only force that acts on them is the weight of each. Hence, at any one place, the *weights of bodies are proportional to their masses.* This is the principle of the common balance. By means of it the weights of bodies are compared and their masses are compared, simultaneously.

42. Centimeter-Gram-Second (C. G. S.) Units.—In all advanced work and research work in Physics throughout the world metric units are used. The centimeter is usually taken as the unit of length, the gram as the unit of mass, and the second as the unit of time. This is called the c.g.s. system of units and the centimeter, the gram, and the second are called the *fundamental units* in the

system, all other units, such as those of force and energy, being *derived* from them, that is, defined by reference to them.

Now it would be convenient, if, along with these three fundamental units, we could use a force equal to the weight of a gram as the unit of force. This, however, is impossible if we are to keep to the equation

$$F = ma \qquad (1)$$

for Newton's second law. For a gram weight acts on a gram mass when it is dropped and it does not produce an acceleration of 1 cm./sec.^{2}, but an acceleration of about 980 cm./sec.^{2} It is evident from (1) that the unit of force must be *the force that, acting on a gram mass, gives it an acceleration of* 1 *cm./sec.*^{2} This is called the *dyne.* The dyne must therefore be about $\frac{1}{980}$ of the gram weight. We also get this result by applying (1) to the case of a falling body, for which (1) becomes

$$w = mg \qquad (2)$$

For if we put $m = 1$, we find that w, the weight of a gram, is equal to g dynes.

A system of units in which the units of length, mass, and time are fundamental units is usually described as an *absolute* system of units, and the preceding system is then called the absolute c.g.s. system; but, for brevity, we shall call it simply the c.g.s. system.

We could also start with the foot, the pound mass, and the second as fundamental units, and define a derived unit of force in a way similar to that in which the dyne is defined. But such a unit is not used (except in some text books). When data are in pounds, feet, and seconds the method of §43 should be used.

43. British Engineering (B. E.) Units.—Engineers are more concerned with weight than with any other force, and (in English speaking countries) they regard the *pound* as meaning a weight, and they take a force equal to a pound as one of the three fundamental units, the other two, in this system, being the foot as unit of length and the second as unit of time. It is then not possible to use the mass of a body that weighs a pound as the unit of mass and at the same time preserve the simple form of equation

$$F = ma \qquad (1)$$

for Newton's second law. For a force equal to a pound weight, acting on a body that weighs a pound, as is the case when the body is allowed to fall, does not produce an acceleration of 1 ft. per sec. per sec. To find how the mass of a body is to be expressed in terms of the fundamental units, we apply (1) to the case of a falling body. This gives

$$w = mg \tag{2}$$

or

$$m = \frac{w}{g} = \frac{\text{weight in pounds}}{g \text{ in ft. per sec.}^2} \tag{3}$$

In taking w/g as the mass of a body, we are, in effect, adopting a new unit of mass, one that is g times greater than the mass of a body that weighs a pound. There is, at present, no generally accepted name for this unit. A name is, however, not essential. It is sufficient to take (3) as giving the *measure*, m, of the mass of a body when (1) is used.

When these British engineering (B. e.) units are used, the equation for Newton's second law and all formulae derived from it *are the same as in the c.g.s. system.* Whenever it is desired, at any stage of solving a problem, to obtain numerical results, the value of m from (3) is substituted for m in the calculations.

In countries that have adopted the metric system of weights and measures, engineers use a system of units that is similar to the preceding, except that the kilogram takes the place of the pound as unit of force, and the meter the place of the foot. We shall not make use of this system in this book. The c.g.s. system of units and the B. e. system will suffice, the former being used when data are in metric units, and the latter when they are in British units.

44. Units in Statics.—The adoption of the dyne as the unit of force in the c.g.s. system and of w/g as the measure of the mass of a body in B. e. units was in order to keep the formula for Newton's second law as simple as possible. But in Statics, whether of solids, liquids, or gases, and also in Heat, the bodies dealt with have no accelerations and Newton's second law is not required. Hence, in these parts of the subject, we may, for simplicity, use the gram force, which is equal to a gram weight, as unit of force in the c.g.s. system (though the dyne is also sometimes used) and the pound mass as unit of mass in the B. e. system.

45. Newton's Second Law (Continued).—The statement of the second law of motion is so brief that some things implied in it might easily escape notice:

1. In the statement of the law, the rate of change of momentum of a body is spoken of, without any reference to whether the body starts from rest or is initially in motion. Hence it is implied that *the effect of a force applied to a body is independent of the state of motion of the body when the force begins to act.* For example, gravity is a force that acts vertically downward. When a body is dropped from a height, the force of gravity gives it a certain acceleration downward; if the same body be started downward with a certain velocity, its acceleration downward will be the same as when the body is simply dropped, and the same will be true if the body be given an initial velocity upward or in any direction. It is found possible to play ball games on a moving steamship; the effect of throwing the ball with a certain force or striking it with a bat is the same as when the steamship is at rest.

2. The law states how a force will affect the motion of a body, but it makes no reference to whether some other force is acting on the body at the same time or not. Hence it is implied that *each force produces its own effect independently of the simultaneous action of any other force;* and, when several forces act on a body, we may calculate the acceleration produced by each as if the other forces did not exist, and then add the accelerations to find the whole effect of all the forces. This very important principle is sometimes called that of the *independence of forces.*

46. Impulse of a Force.—The product of a force and the time during which it acts is called the *impulse* of the force. When a force F acts on a mass m for time t, from the formula for the Second Law of Motion, by multiplying both sides by t, we get:

$$Ft = mat$$

Now at is the increase of velocity produced, and this, multiplied by m, is the increase of momentum. Hence *the impulse of a force equals the momentum produced by it.* If the body, starting with a velocity v_0, has at time t a velocity v,

$$Ft = mv - mv_0$$

47. Newton's Third Law of Motion.—Action and reaction are equal and opposite. In the statements of the first and second

laws of motion, forces acting on bodies are spoken of, but nothing is said as to what exerts force. This lack is supplied by the third law.

The action and reaction referred to here mean *force and counter-force.* The meaning of the statement is that force on any one body is exerted by some other body, and this other body itself experiences an equal and opposite force, exerted by the first body, the line of action of both forces being the line joining the two bodies.

In many cases the truth of this law seems evident. For example, when one presses his two hands against each other, it will be admitted that the hands, if at rest, press equally in opposite directions. If one hand be pressed against a wall, the same must still hold, since the wall merely takes the place of the other hand in the first illustration. But the case is not so clear when a hand is pressed against an obstacle that moves. How, it is sometimes asked, can there be motion produced if the forces are equal and opposite? The answer is that *the two forces spoken of do not act on one body;* there is one force exerted by the hand on the obstacle, and the obstacle yields unless restrained by some other force; the reaction is the back pressure of the body *on the hand,* not a force acting on the body.

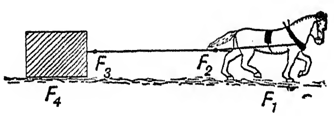


FIG. 16.—Four pairs of actions and reactions.

Consider, also, the forces that come into play when a horse of mass m_1 pulling on a horizontal rope of mass m_2 draws a block of mass m_3. Here there are four pairs of actions and reactions. In the first place, the horse pushes against the ground, and the reaction of the ground is an equal and opposite push. Let the magnitude of the horizontal components of the action and reaction be F_1. Secondly the horse exerts a forward pull, of magnitude say F_2, on the rope, and the reaction of the rope is equal and opposite. The rope exerts a horizontal force on the block, and the block exerts an equal and opposite reaction, the magnitude of each being F_3. Finally, there is the action and reaction between the block and the ground; let the horizontal component of these have a magnitude F_4. If there is an acceleration a, as there must be to begin the motion, F_1, is greater than F_2 by m_1a, F_2 is greater than F_3 by m_2a, and F_3 is greater than F_4 by m_3a. Thus F_1 exceeds F_4 by $(m_1 + m_2 + m_3)a$, and this is, therefore, the total backward push on the ground. When the motion has become constant, $a = 0$ and all the forces mentioned are

of equal magnitude. The reader will find it of interest to consider how the motion of the earth is affected by the force exerted by the horse.

Since a force is always accompanied by a counter-force, the two are parts, or different aspects, of one inseparable whole, and the two together constitute what is called a *stress*. Thus *every force is the partial aspect of some stress*, just as a purchase and a sale are partial aspects of an exchange.

Since the forces between two bodies are equal and opposite, they produce equal and opposite amounts of momentum, and the total momentum of the two bodies remains unchanged. The same principle can be extended to any group of bodies, for their mutual actions consist of pairs of equal and opposite forces. *The total momentum of any mechanical system is not affected by any interaction between its parts.* This is called the principle of the *conservation of momentum.*

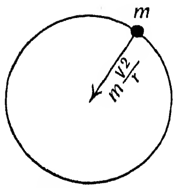


FIG. 17.—A particle moving in a circle is acted on by a force toward the center.

48. Force Required for Motion in a Circle. When a particle revolves in a circle, it has an acceleration toward the center equal to v^2/r (§33), where v is the magnitude of the velocity (*i.e.* the speed) and r is the radius. To cause this acceleration, there must be a force directed toward the center, and, according to Newton's Second Law, this *centripetal* force F must be such that

$$F = m\frac{v^2}{r}.$$

Against this force the particle will exert an equal and opposite reaction on the body that exerts the force toward the center. If, for example, the particle be attached by a string to the finger, the reaction will be a force acting on the finger and will be in a direction outward along the radius. This reaction is called a *centrifugal force.* Thus the centrifugal force is *not a force acting on the moving particle, but a reaction, exerted by the particle, on the other body that exerts the force toward the center.* (That the above formula also applies to the motion of a *body* is shown in §106.)

49. Centrifugal Force.—Illustrations of the above are very numerous, and a few may be mentioned. *Drops of water* are thrown off tangentially from a rapidly moving bicycle or automobile wheel, owing to the fact that there is not a sufficient force toward the center acting on them, and they therefore move off

on a tangent, in accordance with the first law of motion. A *train rounding a curve* presses outward on the rails, and the resultant of this force and the vertical weight of the train is a force inclined to the vertical. Since it is desirable that the whole force should be perpendicular to the sleepers, the outer rail is raised. In the *centrifugal drier*, used in laundries and sugar refineries, the material to be dried is placed in a perforated cylinder, rotating about its axis, which is vertical: the drops of water, not being held by a force directed to the center, escape through the perforations. In the *centrifugal cream-separator*, which is a rotating vertical cylinder, both the milk and the cream tend to move as far from the axis as possible: but the milk, being the denser, exerts the more powerful tendency, and therefore occupies the parts of the vessel farthest from the axis. The *flattening of the earth* at its poles is due to its axial rotation. If at rest, it would be spherical. Being in rotation, it bulges at the equator, until the restoring forces, due to gravitational attractions, supply the requisite force toward the center. The higher the speed of *belting*, the less it presses on a pulley, and the more liable, therefore, is it to slip, for more of the tension of the belting is called on to supply the requisite force toward the center. Watt's *governor for a steam-engine* consists of a pair of balls, whirled around a vertical spindle, at a rate proportional to the speed of the engine. When this speed exceeds the desired limit the outward movement of the balls acts on a steam-valve, so as to decrease the speed of the engine.

RESULTANT OF FORCES. EQUILIBRIUM

50. Resultant of Forces.—Two or more forces may act on a body at the same time. For example, a body that is falling, because of the attraction of the earth, may be blown sidewise by wind pressure. In such cases each force produces an acceleration, independently of the action of the other forces (§45), and the body travels in some path with a definite acceleration, which is the resultant of the accelerations produced by the separate forces.

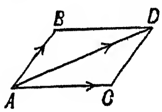


FIG. 18.

The resultant of two or more forces is defined as the single force that would produce the resultant acceleration. The resultant of any number of forces that act on a particle can be found by vector addition, that is, by a triangle, parallelogram, or polygon construction. For a force has a certain magnitude and a certain direction, and is, therefore, a vector quantity. Hence two forces acting on a particle may be represented by lines drawn from a point. Let AB and AC represent two forces F_1 and F_2, acting on a particle. Complete the parallelogram $ABCD$. By the Second Law of Motion the accelerations produced by F_1 and F_2 are in the directions of and proportional to AB and AC, and the resultant acceleration must, therefore, be

represented by AD; and, since the resultant force is the force that will produce the resultant acceleration, it must be in the direction of and proportional to AD. This very important result, called the **Parallelogram of Forces,** is usually stated as follows:

If two forces acting on a particle be represented by two lines drawn from a point and if a parallelogram be drawn with these two lines as sides, the resultant will be represented by the diagonal that passes through the point.

Let θ be the angle between the directions of the forces F_1 and F_2. Then, as in the case of velocities and accelerations,

$$F^2 = F_1^2 + F_2^2 + 2F_1F_2 \cos \theta$$

Since, then, we may add two forces by the parallelogram method or by the triangle method (which is essentially the same), we may in the same way add a third to the resultant of these two and so on. Hence the polygon method of addition applies to forces acting on a particle.

Having found the resultant, F, of the forces that act on a body, we can find the resultant acceleration, a, by the equation in Newton's second law.

51. Resolution of a Force into Components.—Since two or more forces, acting on a particle, can be replaced by a single force, called their resultant, a single force can be replaced by any two or more forces which, added geometrically, give the single force. This is called the resolution of a force into components.

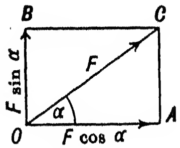


FIG. 19.

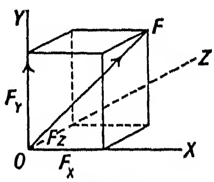


FIG. 20.

The most important case practically is when the components are at right angles to each other. When a single force is resolved into two components (Fig. 19), the components and the force resolved must be in the same plane. When the two components are at right angles, the component that makes an angle α with the whole original force F has a magnitude $F \cos \alpha$, and the other component is $F \sin \alpha$. The agreement as regards the signs of angles noted in §25 applies to the present case.

A force F may also be resolved into three components in three directions at right angles to each other. All that is necessary is to construct a right-angled parallelepiped with the line representing F as diagonal (Fig. 20) and with edges in the three rectangular directions. If the three directions be taken as axes of x, y and z and if the components be denoted by F_x, F_y, F_z respectively, we shall have

$$F^2 = F_x^2 + F_y^2 + F_z^2$$

52. Sliding Down a Smooth Plane.—The force of gravity on a body of mass m acts vertically downward and its magnitude equals mg. If a body is not free to move vertically, but is free to move in some other direction, the only part of gravity that can affect the motion is the component in that direction. For instance, if a body (Fig. 21) is on a *smooth* plane inclined at an angle i to the horizontal, the force of gravity mg may be resolved into a component $mg \sin i$ down the plane and a component, $mg \cos i$ perpendicular to the plane. The latter component will produce pressure on the plane but will not affect the motion down the plane, which will depend only on the former component, $mg \sin i$. If we now apply the formula for Newton's second law to the motion down the plane, we get

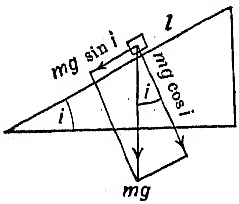


Fig. 21.

$$F = mg \sin i = ma$$

Hence

$$a = g \sin i$$

Let the length of the plane be l, and its height be h. If the body is started from the top of the plane with a speed v_0, and reaches the bottom with a speed v,

$$v^2 = v_0^2 + 2g \sin i \cdot l$$

and, since $l \sin i$ is equal to h

$$v^2 = v_0^2 + 2gh$$

Hence **the speed attained is the same as if the body had fallen vertically through the height of the plane.** This applies only to a body *sliding* down a *smooth* plane. It does not apply to a body

rolling down a plane. If the plane be not perfectly smooth, there will also be a force of friction, say F_f, parallel to the plane, and the resultant force down the plane will be $(mg \sin i - F_f)$.

53. Other Examples of Resolution.—A sail-boat (Fig. 22) effects a resolution of the wind pressure. The wind can exert little force on the sail in a direction parallel to the sail. The force F that it exerts perpendicular to the sail may be resolved into a horizontal component perpendicular to the keel and a component f parallel to the keel. The former produces a small sidewise motion or lee-way, while the latter, being in the direction in which the boat is most free to move, is the effective component.

An airplane creates, by its motion, a "wind" that blows against it and helps to support it, while, at the same time, resisting its motion. The action is very complex, and we shall consider only the effects of the pressure F on the lower side of the wing. This may be resolved into a lifting force, L, and a resisting force or drag, D, that opposes the forward motion. How these vary with the tilt of the wing and with its speed may be inferred from Figure 23. The effect of this wind pressure on the lower side of the wing is, however, of less importance than the effect of the partial vacuum above the wing created by the motion.

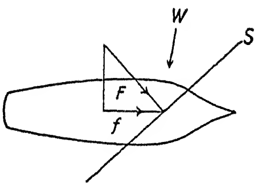


Fig. 22.

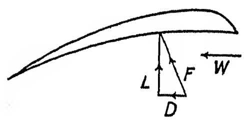


Fig. 23.

54. Analytical Method of Compounding Forces.—A simple and general method of finding the resultant of a number of forces in a plane is to *first resolve each in two directions at right angles, and then add.* Let $F_1, F_2 \ldots$ (Fig. 24) be the forces acting on a particle at O. Take any two convenient rectangular directions, Ox and Oy, and let the angles which $F_1, F_2 \ldots$ make with Ox be $\alpha_1, \alpha_2 \ldots$ respectively. Then F_1 is equivalent to $F_1 \cos \alpha_1$ along Ox and $F_1 \sin \alpha_1$ along Oy, and similarly for the other forces. Let the sum of the components along Ox be denoted by X, and the sum of the components along Oy by Y. Then

$$X = F_1 \cos \alpha_1 + F_2 \cos \alpha_2 + \cdots = \Sigma F \cos \alpha$$
$$Y = F_1 \sin \alpha_1 + F_2 \sin \alpha_2 + \cdots = \Sigma F \sin \alpha$$

We have thus replaced the forces $F_1, F_2 \ldots$ by X along Ox and Y along Oy. The resultant of X and Y is the resultant of $F_1, F_2,$

etc. Let the magnitude of the resultant be R and let it make an angle θ with OX. Then

$$R^2 = X^2 + Y^2$$
$$\tan \theta = \frac{Y}{X}$$

These formulæ give the magnitude and the direction of the resultant. In using this method it must be remembered that, when we substitute for each angle α its numerical value, we must call the angle positive, if it is measured in the direction regarded as positive, say the counterclockwise direction; if measured in the opposite direction it must be regarded as negative.

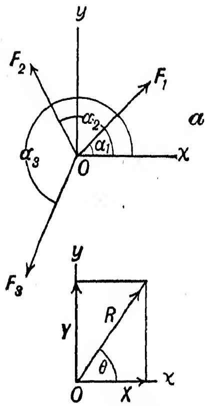


FIG. 24.—Analytical method of compounding forces.

The angle θ that the resultant makes with Ox is found from its tangent. When the tangent is positive, it shows that the angle is between 0° and 90° or between 180° and 270°. To decide between these two, note that the signs of the values of X and Y must be either both positive or both negative, since the tangent is positive. If both are positive, θ is between 0° and 90°; if both are negative, it lies between 180° and 270°. The reader should have no difficulty in completing the reasoning for the case in which $\tan \theta$ is negative.

When X and Y are both zero, that is, when the sum of the components in each of two directions at right angles is zero, R is also zero. Conversely, when R is zero, X and Y must also each be zero, since the square of a number cannot be negative.

When the forces to be compounded are not all in one plane, we may take three directions, Ox, Oy, Oz, at right angles and resolve each force into components in these three directions. Denote the sum of the components along Ox by X, that along Oy by Y, and that along Oz by Z and let the resultant be R. Then

$$R^2 = X^2 + Y^2 + Z^2$$

If $X = 0$, $Y = 0$ and $Z = 0$, then $R = 0$. The converse is also true, since X^2, Y^2 and Z^2 must be either positive or zero.

55. Equilibrium of Forces Acting on a Particle.—*When the resultant of the forces acting on a particle is zero, the forces are said*

to be i equilibrium, that is, in a state of balance, so that they do not change the motion of the particle. The following propositions regarding forces in equilibrium are frequently useful.

1. *When two equal and opposite forces act on a particle, they are in equilibrium*, for their resultant is zero. Conversely, if two forces are in equilibrium, they must be equal and opposite, for otherwise their resultant could not be zero.

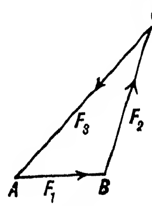


Fig. 25.

2. *When three forces acting on a particle are in the direction of and proportional to the sides of a triangle taken in order, they are in equilibrium.* For if the three forces F_1, F_2, F_3 be in the direction of and proportional to AB, BC, CA, the resultant of F_1 and F_2 will be represented by AC, and the resultant of forces represented by AC and CA is zero.

3. The converse of the last proposition is of very great importance: **If three forces acting on a particle be in equilibrium, and if any triangle be drawn with its sides respectively in the directions of the forces, the forces will be proportional to the sides of this triangle.**

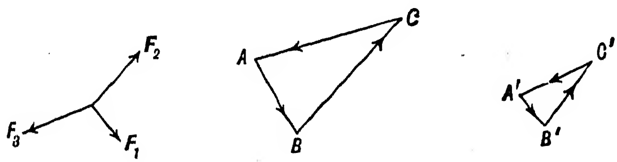


Fig. 26.

To prove this, let us suppose that AB and BC (Fig. 26) are any two lines in the direction of and proportional to two of the forces, F_1 and F_2. Then a force represented by AC is equivalent to F_1 and F_2 taken together. Hence, since the forces are in equilibrium, the third force F_3 must be in the direction of and proportional to CA. Now *any other triangle* such as $A'B'C'$, with sides in the directions of F_1, F_2, F_3 respectively, that is, in the directions of AB, BC, CA, respectively, must be similar to ABC. Hence its sides must be proportional to the sides of ABC, that is, to F_1, F_2, F_3 respectively, which proves the proposition. This converse proposition is important because when we know *the directions* of three forces that are *in equilibrium*, we can find the relative magnitudes of the forces simply by constructing *any* triangle with its sides in the directions of the forces. It should be noticed that the

proposition depends on the principle that two triangles are similar if their corresponding sides are parallel.

4. *When any number of forces, acting on a particle, are in the directions of and proportional to the sides of a closed polygon taken in order, they are in equilibrium;* for the resultant is zero. The converse of this proposition, for more than three forces, is *not* true; for polygons are not necessarily similar when their respective sides are parallel.

5. *When any number of forces are such that the sum of their components in each of three directions at right angles is zero, they are in equilibrium.* This is evident from §54, for, when X, Y and Z are all zero, R must also be zero. Conversely, *when any number of forces are in equilibrium, the sum of their components in any direction equals zero;* for we may take this direction as one of three at right angles, and, since R is zero, the sum of the components in each of these directions is zero (§54).

WORK AND ENERGY

56. Work.—When a force acts on a body, **the product of the force by the distance the body moves in the direction of the force** is, as we shall see later, a very important quantity and is called the *work* performed by the force. Thus, when a force applied to a heavy body raises it a certain vertical distance, work is performed for the force, the amount of the work being the product of the force and the distance of ascent; and, when a horizontal force draws a body horizontally, the work is the product of the force and the horizontal distance.

The phrase "in the direction of the force" that occurs in the definition of work should be carefully noted. When there is no motion in the direction of a force, no work is performed *by that force.* For instance, a travelling crane may, by its chains, exert an enormous force in sustaining a heavy body, and it may move the body through a great distance horizontally, but the force exerted by the chains will do no work if there is no vertical motion.

If a force F acts constantly on a body, while the body moves a distance AB that is not in the direction of the force, to get the work performed we must take the projection of AB on the line of action of the force, and multiply the projection by the force (Fig. 27). If θ is the angle between AB and the direction of the force, the projection, AC, of AB on the line of the action of the force is

$AB \cos \theta$ and the work done is $F \cdot AB \cos \theta$. This at once suggests another method of calculating the work performed, for $F \cdot AB \cos \theta$ is the same as $F \cos \theta \cdot AB$, and $F \cos \theta$ is the component of F in the direction of AB. Thus the work performed by a force is also *the product of the total distance by the component of the force in the direction of motion.*

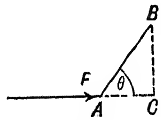


Fig. 27.

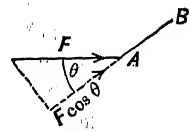


Fig. 28.

Work, or the product of force and distance, must be carefully distinguished from the impulse of a force, which is the product of a force and the time during which it acts (§46). Given a force and the distance, along its line of action, through which it acts, we do not need to know the time in order to calculate the work.

When we say, for brevity, that a force does work, we mean of course, that the body or agent that exerts the force does the work.

57. Positive and Negative Work.—Forces always exist in pairs of equal and opposite forces (§47). Hence, when a force applied to a body does work by moving the body in the direction of the force, it must at the same time overcome an opposing force or reaction. The applied force in this case does positive work, since the motion is in the direction of the applied force. This work is done against the reaction, or we may say that the reaction does negative work, since the motion is in the opposite direction to the reaction.

The nature of the reaction is different in different cases. A horse attached to a wagon is doing work against the force of friction when the wagon is moving uniformly, and the force of friction does negative work. In starting the wagon into motion, the horse does work against the inertia of the wagon, and the reaction is sometimes called an inertia reaction. When a body is moving in some direction and encounters an opposing force, for example when a hammer strikes a nail, the body does positive work against the force, which in this case does negative work.

58. Units of Work.—The unit of work is *the work done by the unit force in acting through unit distance.* When the dyne is taken as unit of force and the cm. as unit of length, the unit of work is that performed by a dyne acting through a cm. and is called an

erg. Since this is a very small unit, a multiple of it, namely 10,000,000 (or 10^7) ergs, is frequently used and is called a **joule.**

When the weight of a pound is taken as unit of force and the foot as unit of length, the unit of work is the work done by a force equal to the weight of one pound when it acts through one foot and is called a *foot-pound.*

59. Diagram of Work.—When a force is constant, to find the work it does we multiply the magnitude of the force by that of the displacement; but, when a force is variable, some other method has to be adopted. One way is to divide the whole displacement up into small parts and multiply each small part by the force at the middle of the small displacement and then add all the products. Stated briefly, $W = \Sigma\ (F \cdot \Delta s)$. By taking the parts small enough, we may get the work as accurately as may be desired. A graphical method is often preferable. It is entirely similar to the method used in finding the distance a point travels when it has a variable velocity (§22). Let OA (Fig. 29) be a line that represents, on some scale, the whole displacement measured in the direction of the force. Divide OA into a very large number of very small equal parts. At O erect a perpendicular OB to represent, on some scale, the force at the beginning of the first part; erect similar perpendiculars to represent the magnitude of the force at the beginning of the other parts, and through the ends of these perpendiculars draw a smooth curve BC. If we calculate the work done in a small displacement ab, by taking for the force its value at the beginning of ab, the result would be too small; and, if we made the calculation by taking the value of the force at the end of ab, the result would be too large and similarly for all the other intervals. By continuing the reasoning as in §22, we find that the actual work done is represented by the area $OBCA$ and can be calculated from the area, allowance being made for the scale in which the drawing is made.

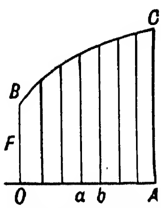


FIG. 29.—Diagram of work.

60. Work in Stretching a Spring.—When the curve of force is a straight line the area may be readily calculated. For example, let us calculate the work done in stretching a spring. In this case it is known that the force that is needed to keep a spring stretched is proportional to the amount of the stretch or increase of length (provided this be not so great as to permanently lengthen the spring). Hence, if the spring is stretched by an amount x, the force applied to it is kx, where k is a constant equal to the force required to produce unit stretch. If, then, a curve be drawn with the values of kx as ordinates

and the values of x as abscissæ, this curve will be a straight line (Fig. 30), which will pass through the origin, since kx is zero when x is zero. Now let the unstretched length be l. To increase the length to $l + x_1$, where x_1 is represented by OL, requires an amount of work represented by the area of OPL or $\frac{1}{2}PL \cdot OL$. Hence the work is $\frac{1}{2}kx_1^2$. Similarly to stretch the spring from length l to length $l + x_2$ requires work $\frac{1}{2}kx_2^2$. Hence the work required to stretch the spring from length $l + x_1$ to length $l + x_2$ is $(\frac{1}{2}kx_2^2 - \frac{1}{2}kx_1^2)$. This is also the work the spring will do in contracting, since, at each step, the force of contraction is equal to the force required to stretch. If the initial stretch be zero, $x_1 = 0$, and the work required to stretch by the amount x is $\frac{1}{2}kx^2$.

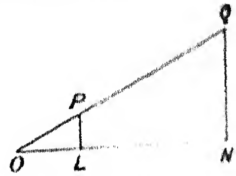


Fig. 30

While we have referred especially to the force exerted by a spiral spring, the above proof and formula evidently apply to the work done by any force that is proportional to displacement. These we shall find later are very numerous.

61. Power or Activity.—The *rate* at which an agent works, or the number of units of work performed per unit time, is called the *power* or activity of the agent. In c.g.s. units the unit of activity is that of an agent that does one erg per second. As this unit is extremely small, the unit employed for most scientific purposes is 10^7 ergs per second, or one joule per second, and is called the *watt*. A still larger unit is the *kilowatt*, which equals one thousand watts.

The unit largely employed for engineering purposes is the *horse-power*, which is the power of an agent that does 550 foot pounds per second or 33,000 foot pounds per minute. If feet be reduced to cms. and weight in pounds to dynes, it will be found that one horse power equals 745.8 watts.

62. Kinetic Energy.—It is often necessary to find the relation between the work done by a force and the velocity of the body to which it is applied, as, for example, in considering the motion of a train. As the simplest case suppose a body of mass m, moving with a velocity v_0 at the beginning of an interval of time t, to be acted on by a single force F in the direction of the velocity, and let the velocity at the end of the time be v and the distance traversed be s. Then from §§28 and 46.

$$s = \frac{1}{2}(v + v_0)t \qquad (1)$$

and

$$Ft = m(v - v_0) \qquad (2)$$

Hence

$$Fs = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2 \qquad (3)$$

One-half the product of the mass of a body and the square of its velocity is called the kinetic energy of the body, or, briefly,

$$K.E. = \tfrac{1}{2}mv^2$$

We may, therefore, state the above conclusion thus:

Work done on body = its gain of kinetic energy;

but it must be remembered this is only for the case in which the force acts on a body which is otherwise free. It is, however, true in all cases, if F stands for the *resultant* of all the forces acting on the body.

We may also reverse the circumstances and inquire what work a body in motion can do, if it meets an opposing force and is brought to rest. Suppose that it exerts a constant force F and does work Fs. Then the opposing force, that is the force applied to the body, will be $-F$. Making this change in (2) we get

$$Fs = \tfrac{1}{2}mv_0^2 - \tfrac{1}{2}mv^2$$

Thus the work done *by the body* against the resistance is equal to the *loss* of kinetic energy of the body.

If the motion continue until the body is brought to rest, v will then be zero. Hence it follows that the *kinetic energy of the body is equal to the work it can do before it is brought to rest.*

It should be noticed that, in the above, v and v_0 stand for the *magnitudes* of the respective velocities, *i.e.*, the speeds (§16). The kinetic energy of a body depends on the square of the magnitude of its velocity and is the same no matter what the direction of motion, that is, kinetic energy is a scalar quantity; to the kinetic energy of one body we may add the kinetic energy of another body and the sum will be the total kinetic energy of both bodies.

Since a force does no work when it is always at right angles to the direction of motion, it follows that, when a body is acted on by a single force at right angles to the direction of motion, the kinetic energy of the body remains constant. Thus, when a body rotates in a circle under the action of a single force directed toward the center, the force does no work and the kinetic energy of the body is constant.

Kinetic energy and work are equivalent quantities; hence the units of kinetic energy are the same as the units of work.

63. Kinetic Energy and Gravity.—The force of gravity on a body is, for small distances above the surface of the earth, a constant force. If a body at a height h_1 above the earth's surface has

a velocity v_1 vertically downward, when it has fallen so that its distance above the surface is h_2, gravity will have done an amount of work, $mg(h_1 - h_2)$; and, if the velocity of the body be then v_2 its kinetic energy will have increased from $\frac{1}{2}mv_1^2$ to $\frac{1}{2}mv_2^2$. Hence

$$mg(h_1 - h_2) = \tfrac{1}{2}mv_2^2 - \tfrac{1}{2}mv_1^2$$

If, on the other hand, the body be projected upward with a velocity v_2 from a height h_2, it will be opposed by the force mg, and the work it will do against gravity, in rising to a height h_1, will be $mg(h_1 - h_2)$. If its velocity at the height h_1 be v_1, its loss of kinetic energy will be $(\frac{1}{2}mv_2^2 - \frac{1}{2}mv_1^2)$. Equating the work done against gravity to the loss of kinetic energy, we get the same equation as before.

In the preceding we have supposed the motion to be vertical; but the result will be unchanged if the motion is not vertical, provided no force except gravity act on the body in the direction of its motion. Any force perpendicular to the motion will do no work and cause no change of kinetic energy. Suppose, for example, the body slides down a smooth plane through a distance l along the plane. Now, we have already shown that it acquires the same velocity as if it fell vertically a distance equal to the height of the plane (§52). The same is true if the descent is along a smooth curve; for a curve may be regarded as made up of very short straight parts to each of which the principle stated will apply. These results are now readily understood by considering the work performed by gravity. For the total displacement in the direction of the whole force of gravity is $(h_1 - h_2)$. Thus the gain of kinetic energy in the descent from the higher level to the lower along any smooth curve must be the same as if the fall were vertical.

64. Kinetic Energy and Elasticity.—When a body is acted on by the force due to a stretched spiral spring, the spring will do work on the body if the spring is contracting, and the body will do work against the force of the spring if it is moving so as to stretch the spring further. Let us first suppose that the body is moving toward the spring with a velocity v_1, the spring being at that moment stretched to an amount x_1 beyond its normal or unstretched length. While the spring is contracting the velocity of the body will be constantly increasing. Let the velocity be v_2 when the spring has contracted so that its stretch is decreased to x_2. In this time (§60) the spring will have done an amount of work $(\frac{1}{2}kx_1^2 - \frac{1}{2}kx_2^2)$ and, since this must equal the increase of kinetic energy of the body,

$$\tfrac{1}{2}kx_1^2 - \tfrac{1}{2}kx_2^2 = \tfrac{1}{2}mv_2^2 - \tfrac{1}{2}mv_1^2$$

We may also suppose the case reversed, that is, we may suppose the body to be moving away from the spring with a velocity v_2 when the stretch of the spring is x_2. Then the velocity of the body will decrease; and if it be v_1 when the stretch of the spring is x_1, work $(\frac{1}{2}kx_1^2 - \frac{1}{2}kx_2^2)$ will have been done against the spring and the decrease of the kinetic energy will be $(\frac{1}{2}mv_2^2 - \frac{1}{2}mv_1^2)$. Equating these we get the same equation as before.

65. Potential Energy.—We shall now consider the two illustrations just given from another point of view. In the case of a body projected vertically upward, there is a loss of kinetic energy equal to mg multiplied by the height of ascent; and, if the body be allowed to descend again, the same amount of work will be performed by gravity and the body will regain its lost kinetic energy. Thus, at the higher level, the body (or rather the body and the earth regarded as one system) has an advantage of position that is equivalent to a certain amount of kinetic energy lost, and this advantage of position is measured by $mg(h_1 - h_2)$. This, since it is equivalent to a certain amount of kinetic energy, is called *potential energy*. Thus it follows that **the sum of the kinetic energy and the potential energy is a constant**, a fact brought out more clearly by writing the equation of §63 thus:

$$\frac{1}{2}mv_1^2 + mgh_1 = \frac{1}{2}mv_2^2 + mgh_2$$

Here, mgh_1 is the increase of the potential energy when the body is raised from the arbitrary zero level (*e.g.*, sea-level) from which h_1 is measured to the height h_1, and a similar statement applies to mgh_2. When the body is at the zero level, it and the earth still possess potential energy, since work could be obtained by allowing the body to fall down a vertical shaft.

The reasoning in the case of a body attached to a spring is so similar to the preceding that it may be left to the reader. The work the spring can do in contracting to its unstretched length is the potential energy of the spring.

The potential energy of a body and the earth depends on their *relative position*, and the potential energy of a spring depends on the relative positions of the parts of the spring. **Potential energy is the capacity a body or system of bodies has for doing work in virtue of the relative positions of its parts.**

We cannot give any universal formula by which potential energy can be calculated, as we can in the case of kinetic energy. In each case we must calculate how much work the body or system

can do in passing from one state to another, and take this as the difference of the potential energy of the body or system in the two states. For any one particular case of potential energy, we may deduce a special expression for its amount, such as those already given for gravity and elasticity.

From the statements made in §§62–65, it is evident that we may define *energy*, of either kind, as *capacity for doing work.*

66. Interchanges of Kinetic and Potential Energy.—We have considered somewhat fully two cases of the interchange of kinetic and potential energy, namely, those of gravity and elasticity, because these are typical and are easily worked out by elementary methods. Such interchanges are common in nature and in industry, and a few may be briefly stated.

(*a*) *Change from Kinetic to Potential.*—When the motion of a train or car is checked by a spring buffer or shock absorber, kinetic energy is changed into potential energy.

As the distance of the earth from the sun increases from midwinter to midsummer, the speed of motion and the kinetic energy decrease, and the potential energy of separation increases.

(*b*) *Change from Potential to Kinetic.*—A clock weight or watch-spring when wound up has potential energy, and this changes to kinetic energy of the pendulum or balance wheel, which would otherwise come to rest.

A bent bow has potential energy, due to the change of position of the particles of the bow and the forces between them. As it unbends it loses this potential energy and the arrow gains kinetic energy.

Water in a lake or above a dam has potential energy; when allowed to escape to a lower level, it loses part of its potential energy and either gains kinetic energy itself or, if it acts on a water wheel or turbine, it imparts kinetic energy to the latter.

(*c*) *Periodic Interchanges.*—In any case of vibration *energy continually changes from the kinetic to the potential form and back again.* Thus, in the vibration of a pendulum, at the bottom of the arc of vibration the potential energy is at a minimum and the kinetic energy is at a maximum, while at the end of the arc of vibration the kinetic energy is zero and the potential energy has increased to a maximum. Similar statements apply to the vibration of a tuning fork, a violin string, a body attached to the end of a wire and vibrating torsionally, the oscillations of the balance-wheel of a watch, and so on.

67. Two Kinds of Forces.—In the preceding we have seen that, when the forces acting between bodies are forces of gravity or forces of elasticity, their action leaves the total kinetic and potential energy of the bodies unchanged, or, as it is usually stated, when only such forces act, the total kinetic and potential energy is *conserved.* Forces whose action between bodies does not cause a change of the total kinetic and potential energy of the bodies are called *conservative forces,* and any system of bodies between which the forces are wholly conservative is called a *conservative system.*

In contrast with these conservative forces stands such a force as friction. A moving body, opposed by friction, loses kinetic energy as its velocity decreases, but it does not at the same time gain potential energy to an equivalent extent. Thus, a body started up a rough inclined plane with a certain velocity will not reach as high a level as it would reach if the plane were smooth, and it will not have as much potential energy when it reaches its highest point. Moreover, its descent will be further opposed by friction, and its store of kinetic and potential energy will thereby be further reduced. Friction, then, is a non-conservative or *dissipative* force since, when in action, it causes a permanent decrease or dissipation of the kinetic and potential energy of a system.

The reason why such a force as gravity has no effect on the sum total of kinetic and potential energy is easily seen. At a certain distance of a body from the earth the force between the two depends only on their distance apart, and is independent of the way in which they are moving. Hence, when they are moving away from each other and are a certain distance apart, they are losing kinetic energy at a rate exactly equal to the rate at which they regain kinetic energy when, at the same distance of separation, they are moving toward one another. Thus forces of gravity between bodies *depend only on the relative positions of the bodies.* The same is true of the forces between the parts of an elastic spring, and this accounts for the fact that such forces of elasticity are also conservative; in fact it is the fundamental characteristic of all conservative forces. But a non-conservative force, such as friction, depends on the way in which a body or a system of bodies is moving; it is always opposed to the direction of relative motion of bodies in contact; hence it causes a diminution of the kinetic energy of the bodies, in whichever direction motion is taking place.

68. The Conservation of Mechanical Energy.—We have seen in the preceding that, *under certain conditions, the total kinetic and potential energy of a system is constant or is conserved.* The conditions referred to are two, (1) the system must not receive energy from or give energy to any outside bodies, (2) the forces between the parts of the system must be wholly conservative. In reality, no system wholly satisfies these conditions. No system is wholly *isolated*, in the sense implied in the first condition; and non-conservative forces, such as friction, are never quite absent. But in many cases these conditions are very nearly satisfied. The solar system, consisting of the sun, planets, and moons, is practically isolated; and, while there are internal frictional forces, such as those of the tides, the work they do is so small compared with the total energy of the system, that their effects in reducing the kinetic energy of the whole have not yet been detected with certainty. Again, the system consisting of the earth and a body vibrating as a pendulum in a vacuum is practically an isolated system free from frictional forces, and the total kinetic and potential energy is very nearly constant; the same is true of a heavy body attached to a spring and vibrating in a vacuum. When, as in cases like these, the conditions are sufficiently nearly satisfied, the principle of the constancy of kinetic and potential energy will often lead to valuable results.

In an isolated system in which there are non-conservative forces, such as friction, energy is expended in doing work against these forces; and if, to the sum of the kinetic and potential energy, we add the work done against non-conservative forces, the sum will be constant. But what becomes of the energy so expended? For long it was supposed to be wholly lost. It was, of course, known that heat was produced when work was done against friction; but heat was supposed to be a form of matter. But about 1840 the view was advanced that heat, instead of being a form of matter, is energy of the particles of a body, and this led to the discovery of the **Law of Conservation of Energy,** which is considered later under "Heat."

ROTATION

69. Angular Displacements.—In §§9–35, we studied the motion of translation of a point, as a preliminary to the study of the effect of forces on the motion of particles and of bodies moving without rotation. We shall now consider the motion of bodies in rotation,

as a preliminary to studying the effects of forces on the motion of rotation of bodies.

The motion of a body is one of rotation when each point in the body moves in a circle, the center of which is on a straight line called the axis of rotation. All points in the body turn in any time through equal angles, and the angle described in any time is called the *angular displacement* of the body in that time. Its magnitude may be stated in degrees or in radians (1 radian = 57.3° approx.), but the latter method is in many ways the more convenient for the present purposes.

70. Angular Velocity.—The rate of rotation of a body is called its *angular velocity*. When the angular displacements of a body in all equal times are equal, the velocity is a constant angular velocity, and the magnitude of the angular velocity is *the angle through which the body turns in unit time.* If the angle is reckoned in radians and the second is taken as unit of time, the magnitude of the angular velocity is the number of radians described in one second. The unit of angular velocity is *one radian per second.*

If the velocity is not constant, as, for example, when a fly-wheel is being set in motion or stopped, the angular velocity or rate of angular displacement is defined in the same way as in the analogous case of variable linear velocity (§19), that is to say, we must take the average angular velocity in a short time, and then suppose this time indefinitely decreased, so that the average angular velocity approaches a limiting value, which is the *instantaneous angular velocity.*

71. Angular Acceleration.—The rate of increase of the angular velocity of a body is called its *angular acceleration.* When the angular velocity increases by equal amounts in equal times, the angular acceleration is constant and its magnitude is the *increase of angular velocity in unit time.* If we denote the angular acceleration by α, the increase of angular velocity in each second is α, and the increase in t seconds is αt. Hence, if, at the beginning of an interval of time t, the angular velocity is ω_0, and at the end of the interval it is ω,

$$\omega = \omega_0 + \alpha t \tag{1}$$

In this time the body has turned through a certain angle, say ϕ. To find the magnitude of ϕ, we may represent the varying values of the angular velocity by means of a curve of angular velocity,

as we did in the similar case of a varying linear velocity (§28), and the area of the diagram will represent the angle ϕ. The two diagrams would have precisely similar properties, the only difference being that in one case we would speak of linear displacement, s, linear velocity, v, linear acceleration, a, while in the other case we would speak of angular displacement, ϕ, angular velocity, ω, and angular acceleration, α. Hence, when the angular acceleration is constant, the formula for ϕ, which must be precisely similar to (2) of §28, is

$$\phi = \omega_0 t + \tfrac{1}{2}\alpha t^2 \tag{2}$$

By elimination of t between (1) and (2) we get

$$\omega^2 = \omega_0^2 + 2\alpha\phi \tag{3}$$

Also, as in the similar case of (4) of §28, the angle of rotation can be stated in terms of the mean angular velocity

$$\phi = \frac{\omega + \omega_0}{2} t \tag{4}$$

72. Angular Velocity and Linear Velocity.—When a point revolves at a constant rate in a circle, its motion may be described either by means of its angular velocity, ω, or by its linear velocity, v, along the tangent, and there is a simple relation between the two. Let the radius of the circle be r and let the time in which the point moves from P to Q be t. Denoting the length of the arc PQ by s and the angle POQ by ϕ, we have, from the definitions of linear and of angular velocity,

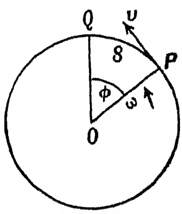


FIG. 31.

$$s = vt. \quad \phi = \omega t.$$

Now in radian measurement $\phi = s/r$. Substituting in this the values of s and ϕ, we get

$$\omega = v/r$$

Thus the relation between angular velocity and linear velocity, when a point rotates in a circle, is the same as the relation between an angle and the arc which it subtends.

This relation can be obtained from $\phi = s/r$ by differentiation with reference to t, r being constant

More briefly stated, the proof of the last, very important, relation amounts to this: v is the length of arc described per second; hence v/r is the angle described per second in radian measurement, that is, the angular velocity.

When a point describes a circle with variable speed, the same relation holds true, with the understanding that ω and v are the instantaneous values of the angular and the linear velocity respectively. The proof is the same as before, t being taken as a very short interval.

When a body rotates about an axis with angular velocity ω, a point in the body describes a circle of radius r, and r is different for points at different distances from the axis. If r and r' are the respective distances of two points from the axis and v and v' their respective linear velocities, $v = r\omega$ and $v' = r'\omega$. Hence $v:v'::r:r'$.

73. Instantaneous Axes of Rotation.—When the axis about which a body rotates varies from moment to moment, the preceding applies to the values of ω, v, and r at any moment. For example, the wheel of a moving wagon or bicycle is always in contact with the road, and the point of contact is, at any moment, the point about which the whole wheel is rotating at that moment. Now the top of the wheel is twice as far from the ground as the center of the hub, and it must, therefore, have twice as great a linear velocity.

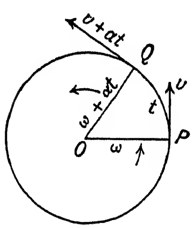


Fig. 32.

74. Angular Acceleration and Linear Acceleration.—When a point revolves in a circle (Fig. 32) with changing angular velocity, it has an angular acceleration, say α. The speed of the point along the tangent increases with an acceleration, say a. The same relation holds between a and α as between v and ω (§72). For, if ω is the angular velocity at the beginning of a short time t, and v the linear speed at this time, $v = \omega r$; at the end of the time t the angular velocity is $(\omega + \alpha t)$ and the linear speed is $(v + at)$. Hence $(v + at) = r\ (\omega + \alpha t)$. From these two equations we get

$$a = r\alpha$$

More briefly stated, a is the added linear speed per unit time, and α the added angular velocity per unit time, and the relation between angular velocity and linear speed must hold true of these increases.

This can also be obtained from $\omega = v/r$ by differentiation with reference to t.

It should be noted that a here means the *rate of change of speed along the tangent.* Since the direction of the velocity is also changing, this cannot be the only acceleration. In fact, as we have already seen (§33), there is, in all cases of motion in a curve, a linear acceleration toward the center, equal to v^2/r, or, as we may now write it, $\omega^2 r$, since $v = r\omega$.

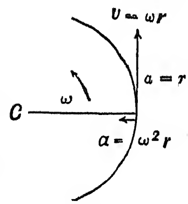


Fig. 33.

These relations, which are very important, are summarized in Fig. 33.

75. Graphical Representation of Angular Quantities.—An angular displacement is of a certain magnitude and is about a certain axis. Given the axis, the direction of rotation around it, and the magnitude of the angular displacement, we know everything about it. Now, all these can be represented graphically by *a length marked off on the axis,* so as to represent, to some scale, (*e.g.*, a cm. per radian) the magnitude of the angular displacement. There must also be some agreement as to which direction along the axis shall represent a certain direction of rotation about the axis. The rule usually adopted for this purpose is called the "right-hand screw rule," namely, **the direction of the line that represents an angular displacement is related to the direction of the rotation as the direction of translation is to the direction of rotation of an ordinary (right-hand) screw.** For example, a line to represent the angular displacement of the earth in 24 hours, due to its rotation about its axis, would be drawn from the center toward the N. pole. Two lines to represent the angular displacements of the hands of a watch in one hour would be drawn through the center of the face toward the back, and the one for the minute hand would be twelve times as long as the one for the hour hand.

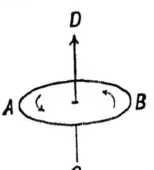


Fig. 34.—A rotation indicated by the arrows is represented by CD.

A line to represent an angular velocity would be laid off on the axis of rotation according to the rule stated, and a line to represent an angular acceleration would be drawn in the same way.

A directed line that represents an angular displacement according to the agreement stated is a vector, since it has both magnitude and direction; but it differs from vectors that represent linear displacements in the fact that it must, in any diagram, be located on a certain line, namely, the line that stands for the axis of rotation.

Such a vector is therefore called a *localized vector* or *rotor*. Two parallel and equal vectors of this kind, not in the same line, do not represent the same angular displacement, since the rotations they represent are about different axes.

76. Addition of Angular Velocities and Accelerations about Intersecting Axes.—A body may have two or more simultaneous angular velocities. For example, suppose a bicycle wheel, while rotating about its axis, is mounted on a horizontal platform, which is kept in rotation about a vertical axis. At any moment, the wheel has two component angular velocities about intersecting axes. Each may be represented by a vector drawn from the center according to the rule stated in §75. We may then add these two vectors by the parallelogram method, and the diagonal will present, in magnitude and direction, the resultant angular velocity at the moment in question.

Since angular accelerations are increments of angular velocities per unit time, we may add them as we add angular velocities.

CENTER OF MASS

77. General Description of Center of Mass.—When the motion of a rigid body is one of translation, without rotation, all points in the body move in exactly the same way, and, in describing or calculating the motion, any point in the body may be taken as representing the whole body. When the motion is one of translation combined with rotation, different points in the body move differently, and there is no one point the motion of which completely represents the motion of the whole body. There is, however, in any body, one particular point that, for many purposes, may be taken as representing the body, so that, for these purposes, the body may be regarded as concentrated to a particle at that point. This point, which we shall define presently, is the *center of mass* of the body. For instance, let a uniform circular disk be tossed into the air; it will be seen that the center of the disk moves like a particle, either in a straight line or in a parabola, while other points in the disk rotate around it. If the disk is loaded with lead on one side, it will be some other point, not the geometrical center, that will possess this property.

If a body were wholly free and were struck a blow at random, it would start with both translation and rotation; but, if the blow

were applied at the center of mass or in a line through the center of mass, the motion would be one of translation, without rotation.

The center of mass is thus seen to be a point of great importance in describing or calculating the whole motion of a body. In what follows we shall define the center of mass and show how its position may be calculated. Then, from the definition, we shall deduce the above and other properties.

78. Center of Mass of a Number of Particles.—The meaning of the center of mass, in general, will be more clearly understood if we begin with some simple cases that will suggest the general definition.

(1) *Two Particles.* Let the particles be m_1 at P_1 and m_2 at P_2 (Fig. 35). Let C_1 be a point that divides P_1P_2 inversely as the masses of the particles, that is, such that

$$m_1 \cdot C_1P_1 = m_2 \cdot C_1P_2$$

C_1 is the center of mass of m_1 and m_2.

FIG. 35.

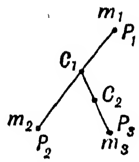


FIG. 36.

(2) *Three Particles.* Let the particles be m_1 and m_2 as above and m_3 at P_3 (Fig. 36). Suppose m_1 and m_2 to be replaced by $(m_1 + m_2)$ at C_1, and let C_2 be a point in C_1P_3 such that

$$(m_1 + m_2) \cdot C_1C_2 = m_3 \cdot C_2P_3$$

C_2 is the center of mass of m_1, m_2 and m_3.

(3) *Any Number of Particles.* Proceeding as above, we get the center of mass, C, of any number of particles, and the same will apply to a body of any form, since it may be divided up into a large number of small parts.

We shall show in the next section that the point to which such a process leads is independent of the order in which the particles are taken.

79. Distance of Center of Mass from a Plane.—Let EF (Fig. 37) be the straight line in which any plane is cut by a perpendicular plane through P_1P_2 of

§78 (1). Draw P_1L, P_2M, C_1N perpendicular to EF and denote their lengths by d_1, d_2 and D_1, respectively. Draw P_1Q_1 and P_2Q_2 perpendicular to C_1N. Since C_1Q_1 and C_1Q_2 are the projections of C_1P_1 and C_1P_2, it is readily seen, from the equation in §78 (1), that

$$m_1 \cdot C_1Q_1 = m_2 \cdot C_1Q_2$$

or

$$m_1(d_1 - D_1) = m_2(D_1 - d_2)$$

$$\therefore (m_1 + m_2)D_1 = m_1d_1 + m_2d_2$$

If we should proceed to apply the same method to $(m_1 + m_2)$ at C_1 and m_3 at P_3 (Fig. 38), we would, it is evident, get a similar result. By extending the same method to any number of particles we shall obtain the general formula

$$(m_1 + m_2 + \cdots)D = m_1d_1 + m_2d_2 + \cdots$$

where D is the distance of the center of mass of all the particles from the plane.

It is evident that this result will not be altered if the order in which the various particles are taken is altered in any way, and that it is true whatever plane of reference is chosen.

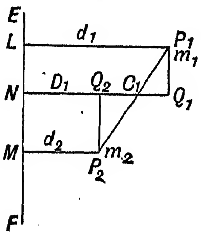


FIG. 37.

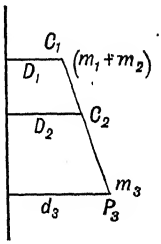


FIG. 38.

80. General Definition of Center of Mass.—If $m_1, m_2, \ldots$ are the respective masses of the particles constituting a body (or group of particles) of total mass M, and if the respective distances of these particles from any plane are $d_1, d_2, \ldots$, the center of mass is a point whose distance from the plane is

$$D = \frac{m_1d_1 + m_2d_2 + \cdots}{m_1 + m_2 + \cdots} = \frac{\Sigma md}{M}$$

If, in any case, one or more of the distances are measured on the opposite side of the plane from the others, in substituting numbers for the various distances, those corresponding to one side of the plane must be given positive signs and the others negative.

From this general definition of center of mass, the formulae for the center of mass of two or three particles in §78 can readily be derived.

If the plane from which d_1, d_2 . . . are measured passes through the center of mass, D is zero and in this case

$$m_1d_1 + m_2d_2 + \cdots = 0$$

When it is desired to find the position of the center of mass of a body by applying the above formula, it is only necessary to apply it to distances from three planes at right angles. Denoting the distances from one of them by x's, from a second by y's, and from the third by z's, we get

$$\bar{x} = \frac{\Sigma mx}{M} \qquad \bar{y} = \frac{\Sigma my}{M} \qquad \bar{z} = \frac{\Sigma mz}{M},$$

where $\bar{x}$ denotes the distance of the center of mass from the plane from which the x's are measured, and similarly for $\bar{y}$ and $\bar{z}$.

81. Center of Mass of a Regular Body.—The center of mass of two equal particles is at the middle of the line joining them. A uniform rod may be divided into pairs of equal particles, the two in each pair being equidistant from the center of the rod. Hence the center of mass of the whole rod is at its middle point. Similar reasoning may be applied to any homogeneous body which has a geometrical center, such as a circle, ellipse, sphere, spheroid, parallelogram, cube, parallelopiped, etc. The center of mass of each of these is at its geometrical center.

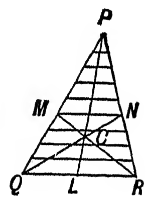


Fig. 39.

When a body can be divided into parts such that the center of mass of each is known, the center of mass of the whole can usually be found. A triangle may be divided into narrow strips parallel to one side; the center of mass of each strip lies on the line joining the middle of that side to the opposite vertex. Hence the center of mass of a triangle is at the intersection of the three lines which join the vertices to the middle of the opposite sides. Similar reasoning shows that the center of mass of a triangular pyramid is at the intersection of the four lines that join the vertices to the respective centers of mass of the opposite faces.

82. Velocity and Acceleration of the Center of Mass.—Let us suppose that the velocity of each particle in a group of particles is known. How can the velocity of the center of mass be found? To answer this it is sufficient to show how the velocity of the

center of mass in a direction perpendicular to each of three planes at right angles can be found.

To find the velocity of the center of mass in a direction perpendicular to any plane, consider the distances of the particles and of the center of mass from that plane. These are connected by the equation (§80)

$$(m_1 + m_2 + \cdot \cdot \cdot)D = m_1d_1 + m_2d_2 + \cdot \cdot \cdot \qquad (1)$$

At a time t later these distances will have all changed. Let the new values of the distances be d_1', d_2' . . . D'. Then

$$(m_1 + m_2 + \cdot \cdot \cdot)D' = m_1d_1' + m_2d_2' + \cdot \cdot \cdot \qquad (2)$$

Subtract each side of (1) from the corresponding side of (2); divide through by t and suppose t decreased without limit.

$$(D' - D)/t$$

will become the velocity, say $\bar{v}$, of the center of mass;

$$(d_1' - d_1)/t$$

will become the velocity, say v_1, of m_1 and so on. Hence

$$(m_1 + m_2 + \cdot \cdot \cdot)\bar{v} = m_1v_1 + m_2v_2 + \cdot \cdot \cdot \qquad (3)$$

Thus the velocity of the center of mass perpendicular to any plane is related to the velocities of the separate particles in that direction as the distance of the center of mass from the plane is related to the distances of the particles from it.

We may now proceed to apply the same reasoning to find the acceleration of the center of mass. Starting with (3) above, let us consider what (3) becomes at a short time t later. We shal thus get two equations. Subtracting one from the other, as before, dividing by t, and then supposing t indefinitely short, we get

$$(m_1 + m_2 + \cdot \cdot \cdot)\bar{a} = m_1a_1 + m_2a_2 + \cdot \cdot \cdot \qquad (4$$

Equation (3) is readily obtained by differentiating (1) with reference to the time (see §19) and (4) is obtained by differentiating (3) (see §32).

83. Acceleration of Center of Mass Due to External Forces.—Equation (4) of the last section has a very important interpretation. The term m_1a_1 is, by the Second Law of Motion, equal to the force that acts on m_1 in the direction in which a_1 is measured

which, of course, may be any direction, and similarly for the other particles. Now the forces may be divided into two groups, (1) forces applied from the outside, or *external* forces, such as gravity acting on the body, pressures and pulls applied to the surface of the body, and so on; (2) forces that the particles exert on one another, that is, *internal forces*, actions and reactions between the particles. By the Third Law of Motion these internal forces occur in pairs of equal and opposite forces, and the sum of the components of all of them in any direction is zero.

Hence the right side of (4) stands for the sum of the components, in the direction considered, of all the *external* forces. Thus, if M be the whole mass of the body or group of particles,

$$\bar{a} = \frac{\textit{sum of components of external forces}}{M}$$

Now, by the Second Law of Motion, this is the expression we would arrive at, if we asked, "what acceleration would the center of mass of the body receive, if the whole mass were concentrated there, and all the external forces were transferred, parallel to themselves, so as to act at that point?"

Hence **the center of mass of a body moves as if the whole mass were concentrated at the center of mass and the forces acting on the body were transferred, with their directions unchanged, to the center of mass.**

We now see the explanation of the facts stated in §77. In the case of a body tossed into the air, gravity is the only external force, and the center of mass moves as if all the mass and weight were concentrated there, that is, it moves as a particle would. Even when a body has its form changed very abruptly by the action of internal forces, as in the case of the explosion of a rocket, the internal forces do not affect the motion of the center of mass of all the particles. When two bodies approach and impinge, the motion of their center of mass is not affected by the forces between the bodies during impact, and hence continues unchanged after the impact. There are powerful forces of attraction between the sun and the planets that make up our solar system, but these forces do not have any effect on the motion of the whole system through space.

84. Translation and Rotation.—It is evident that, to ascertain the whole motion of a body, it is sufficient to find: (1) The motion

of translation of some point in the body, (2) The motion of rotation about that point. By the result obtained in §83 we can find the linear acceleration of the center of mass from the magnitudes and directions of the forces, without considering the distances of their lines of action from the center of mass.

We shall now consider how the angular acceleration of a body can be calculated, but, in an elementary work, it is necessary to confine attention, for the most part, to the rotation of rigid bodies mounted on fixed axes.

MOMENT OF FORCE AND ROTATIONAL INERTIA

85. When a rigid body is mounted on a fixed axis (*e.g.* a grindstone or fly-wheel), the only motion that a force applied to it can produce is one of rotation about the axis. To find the magnitude of the effect, we must consider, not only the magnitude and the direction of the force, but also the distance of its line of action from the axis. For it is a matter of common experience that a force can be most effectively applied to set a large body into rotation when it is applied as far from the axis as possible.

On the other hand, the inertia resistance that the force encounters depends on something more than the mass of the body. For it is also well known that the farther, on the whole, the mass of the body is from the axis, as, for example, in the case of a fly-wheel with a heavy rim and light spokes, the harder it is to set it into rotation or to stop it.

We are thus led to consider moments of force and rotational inertia.

86. The Moment of a Force.—Consider a body mounted on a fixed axis, A, perpendicular to the plane of the paper (Fig. 40). Let a force, f, act on the body, the line of action of the force being in a plane perpendicular to the axis, and let the perpendicular distance of the line of action from the axis be p. The product fp is called the moment of f about A. It depends on the magnitude, direction, and line of action of the force; but it does not depend on the particular point in the line of action at which f is applied. A moment of force is also called a *torque*.

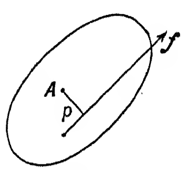


Fig. 40.

The above is not a general definition of the moment of a force, for we have supposed the line of action of the force to be in a plane

perpendicular to the axis. To find the moment of a force, F, in any direction, we must first resolve F into a component parallel to the axis and a component, say f, perpendicular to the axis. The former cannot produce motion about the axis, since it is parallel to the axis. The latter component, f, is the effective component.

The moment of a force about an axis is the product of the component of the force perpendicular to the axis (the other component being parallel to the axis) by the perpendicular distance of this component from the axis.

Since one direction of rotation about an axis is taken as positive, the other being taken as negative, moments of forces are considered as positive or negative according to the directions in which they tend to produce rotation.

A moment of force, although it is the product of two quantities f and p, should be thought of as a single physical quantity, just as work, the product of F and s, is a single physical quantity. As such we shall denote it by L.

87. Work Done by a Moment of Force.—Let the line of action of the force be fixed relatively to the body, and let the moment of force about the axis be constant and equal to fp. When the body rotates through an angle θ in the direction of the moment, the force acts through a distance $p\theta$, and the work done is $fp\theta$ or $L\theta$. Hence in rotation,

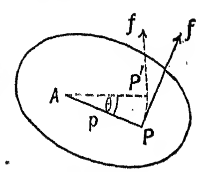


FIG. 41.

Work = moment of force × angular displacement

The similarity of this expression to that for the work done in translation, namely Fs, should be noted, moment of force corresponding to force and angular displacement to linear displacement.

88. Kinetic Energy of Rotation.—Each particle of a rotating body has a certain linear velocity and a certain amount of kinetic energy, and the total kinetic energy is the sum of the kinetic energies of the particles. A particle of mass m, at a distance r from the axis, has a linear velocity ωr, where ω is the angular velocity; and the kinetic energy of the particle is $\frac{1}{2}m\omega^2r^2$. Now r is different for different particles, but ω is the same for all. Hence

$$E = \tfrac{1}{2}(m_1r_1^2 + m_2r_2^2 + \cdot\cdot\cdot)\omega^2$$

The term in brackets depends only on the mass and the form of

the body and on the particular axis of rotation considered. If we denote it by I,

$$E = \tfrac{1}{2}I\omega^2$$

an expression similar to that for kinetic energy of translation, namely $\frac{1}{2}mv^2$.

89. Rotational Inertia.—The expression denoted above by I is called the *rotational inertia* or *moment of inertia* of the body about the particular axis of rotation. It may be defined as *the sum of the products of the particles by the squares of their respective distances from the axis of rotation,* or briefly,

$$I = \Sigma(mr^2).$$

The meaning of this expression for I may seem clearer, if it is stated in a different way. Let the mass of the body be M and

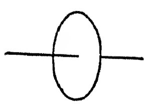


FIG. 42.

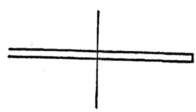


FIG. 43.

suppose it to be divided into a large number N of small parts, each of mass M/N. Then

$$I = M\frac{\Sigma r^2}{N} = Mk^2$$

where k^2 is the mean of all the values of r^2 and k is called the *radius of gyration* about the axis. All parts of a thin hoop, of mass M and radius a, may be regarded as being at the same distance from its geometrical axis, and its rotational inertia about that axis is, therefore, Ma^2. Two other important cases are the following: For a circular disk of radius a about its geometrical axis (Fig. 42)

$$I = \tfrac{1}{2}Ma^2$$

For a thin rod of length l about a transverse axis through its center (Fig. 43)

$$I = \tfrac{1}{12}Ml^2$$

The derivation of such formulæ is best performed by means of the Integral Calculus.

If ρ be the mass per unit area of the disk, and if it be supposed divided into hoops, I is the integral from 0 to a of $2\pi r dr \rho r^2$ which is $\frac{1}{2}\pi\rho a^4$ or $\frac{1}{2}Ma^2$.

If ρ be the mass of unit of length of the rod, I is the integral from $-l/2$ to $l/2$ of $\rho r^2 dr$ which is $\frac{1}{12}\rho l^3$ or $\frac{1}{12}Ml^2$.

90. Energy Equation for Rotation.—Consider a rigid body on a fixed axis and acted on by a moment of force L, the rotational inertia about the axis being I. If the body turn through an angle ϕ the work done will be $L\phi$. Since the body is rigid, the relative positions of its particles will not be changed, and there will, therefore, be no change in their potential energy. Hence the work done will equal the increase of kinetic energy, or

$$L\phi = \tfrac{1}{2}I(\omega^2 - \omega_0{}^2)$$

91. Angular Momentum.—In the case just considered, let the time of rotation through the angle ϕ be t. The average angular velocity in the time t is $\frac{1}{2}(\omega + \omega_0)$. Hence

$$\phi = \tfrac{1}{2}(\omega + \omega_0)t$$

From this and the equation of §90 we get

$$Lt = I(\omega - \omega_0)$$

The product Lt evidently corresponds to the product Ft in the case of translation (§46) and may be called the impulse of the moment of force. The expression on the right is the increase of $I\omega$, which corresponds to momentum, mv. From this analogy the *product of rotational inertia and angular velocity* is called *angular momentum.*

If $L = 0$, that is, if the body is not acted on by any force having a moment about the axis of rotation, $I\omega = I\omega_0$, or the angular momentum is constant.

92. Conservation of Angular Momentum.—The last statement is a particular case of the principle called the conservation of angular momentum, namely, the case of a rigid body mounted on a fixed axis. A wobbling quoit is a more general case. If we neglect air resistance, the only external force is gravity, which acts through the center of mass. The angular momentum is constant in amount, but its axis has a periodic motion. An acrobat turning in the air and the projectile from a rifled gun are other illustrations. The whole solar system illustrates the general principle, which is that *the total angular momentum of any system of bodies, not acted on by forces*

having a resultant moment about the center of mass, is constant in amount.

93. Angular Acceleration Produced by a Moment of Force.—From §91 it follows that

$$L = I\frac{\omega - \omega_0}{t}$$

Hence if α be the angular acceleration produced by L,

$$L = I\alpha$$

This is the fundamental equation for calculating the motion of a body on a fixed axis. It corresponds to $F = ma$ for translation.

If the body is not on a fixed axis, but is free, the applied force will produce linear acceleration of the center of mass (§83) and also rotation about some axis through the center of mass, the direction of which depends on the shape of the body. If the line of action of the force is in the plane of two of the principal axes (§106), the rotation will be about the third, but the subject cannot be discussed fully here.

94. Rotational Inertias about Parallel Axes.—There is a simple and useful relation between the rotational inertia, I, of a body about any axis and its rotational inertia, I_0, about a parallel axis through the center of mass, namely,

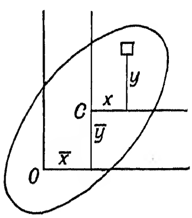


FIG. 44.

$$I = I_0 + Mh^2,$$

M being the mass of the body and h the distance between the two axes. The proof of this is as follows:

Let the axes referred to be perpendicular to the plane of Fig. 44 and cut the plane in O and C respectively. Let the coördinates of a particle, m, referred to rectangular axes through C be x and y and let the coördinates of C referred to a parallel set of axes through O be $\bar{x}$ and $\bar{y}$. Then

$$I = \Sigma m\{(x + \bar{x})^2 + (y + \bar{y})^2\}$$
$$I_0 + Mh^2 = \Sigma m(x^2 + y^2) + \Sigma m(\bar{x}^2 + \bar{y}^2)$$

Let us now expand the right-hand sides and compare them, term by term. It will be found that they differ only as regards $2\bar{x}\Sigma mx$ and $2\bar{y}\Sigma my$. Now $\bar{x}$ and $\bar{y}$ are constants. Also x and y are dis-

tances from planes through the center of mass, and therefore $\Sigma mx = 0$, and $\Sigma my = 0$ (§80). Hence I and $I_0 + Mh^2$ are equal.

95. Kinetic Energy of a Body which has Translation and Rotation.—Let the body of Fig. 44 be in rotation about the axis through O, supposed fixed in the body (or in fixed connection with it), with angular velocity ω. This is also its angular velocity about the axis through C, since both revolutions are completed in the same time. The total kinetic energy is $\frac{1}{2}I\omega^2$. From the equation of §94,

$$\begin{aligned}\tfrac{1}{2}I\omega^2 &= \tfrac{1}{2}I_0\omega^2 + \tfrac{1}{2}Mh^2\omega^2 \\ &= \tfrac{1}{2}I_0\omega^2 + \tfrac{1}{2}MV^2\end{aligned}$$

since V, the linear velocity of C, is equal to $h\omega$. Thus the total kinetic energy may be regarded as consisting of

(1) K. E. of translation of the c. of m.
(2) K. E. of rotation about the c. of m.

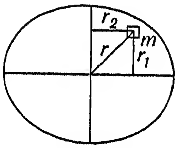


Fig. 45.

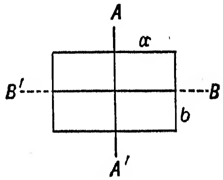


Fig. 46.

This applies to a cylinder rolling down a plane or a moving carriage wheel, since there is an axis which is, for the moment, at rest, namely, the line of contact with the surface. It holds, likewise, for any body rotating about an axis through the center of mass and also moving perpendicular to that axis, for there is in all such cases an instantaneous axis (§73). In fact, by resolving the linear velocity parallel to and perpendicular to the axis, it will be seen that the principle is true for any motion of a rigid body.

96. Rotational Inertia of a Disk.—If the rotational inertia of a disk of any shape about two axes at right angles in the plane of the disk be I_1 and I_2, its rotational inertia about a third axis intersecting the two and perpendicular to the plane of the disk is $I_1 + I_2$.

For let the distances of an element, m, of the disk from the first two axes be r_1 and r_2 respectively (Fig. 45).

$$\begin{aligned} r^2 &= r_1^2 + r_2^2 \\ mr^2 &= mr_1^2 + mr_2^2 \end{aligned}$$

Summing up for all elements

$$I = I_1 + I_2.$$

97. Some Other Rotational Inertias.—A uniform rectangular disk may be divided into rods and their rotational inertias added (Fig. 46). Thus about axes in the plane of the disk and bisecting pairs of opposite sides $I_a = \frac{1}{12}Ma^2$, $I_b = \frac{1}{12}Mb^2$. Hence by §96 the rotational inertia of the disk about an axis through the center perpendicular to the disk is

$$I = \frac{1}{12}M(a^2 + b^2)$$

A uniform rectangular block may be divided into rectangular disks. Hence its rotational inertia about an axis through its center and perpendicular to a face with sides equal to a and b is given by the last formula.

The rotational inertia of a uniform circular disk about any two axes in the plane of the disk that pass through the center and are at right angles are equal. Hence by §96, each is equal to $\frac{1}{4}Ma^2$.

A circular cylinder may be divided into circular disks. Hence its rotational inertia about its geometrical axis is $\frac{1}{2}Ma^2$. Its rotational inertia about a transverse axis through its center may also be found by the above principles, but we leave it as an exercise.

98. Units in Rotation.—The formulæ we have been using since Newton's second law was stated (§41) are valid for both c.g.s. and B.e. units (see §43). For example, $I = \Sigma mr^2$, $I(\text{disk}) = \frac{1}{2}Ma^2$, $L = I\alpha$, etc. are the same in both sets of units. In using c. g. s. units, m or M is in grams. When B.e. units are used and numerical results are to be obtained by substituting numbers for symbols, M is to be calculated from $M = w$ (lbs) $\div$ g (ft/sec.^{2}).

TABLE OF ROTATIONAL INERTIAS

Body	Axis	Rotational Inertia
Rod	transverse through end	$\frac{1}{3}Ml^2$
Rod	transverse through middle	$\frac{1}{12}Ml^2$
Circular disk	perpendicular through center	$\frac{1}{2}Ma^2$
Circular cylinder	longitudinal through center	$\frac{1}{2}Ma^2$
Circular cylinder	transverse through center	$M(\frac{1}{4}a^2 + \frac{1}{12}l^2)$
Rectangular block	through center perpendicular to face with sides a and b in length	$\frac{1}{12}M(a^2 + b^2)$
Sphere	through center	$\frac{2}{5}Ma^2$

RESULTANT OF FORCES ACTING ON A BODY

99. Resultant.—When treating of the forces acting on a *particle*, we found that they could always be replaced by a single equivalent force, called their resultant. When a number of forces act on a *body*, they are in certain cases equivalent in their effects

to a single force, which is called their resultant. As we shall see later, there are other cases in which this is not so.

100. Conditions to be Satisfied by Resultant.—1. The resultant must be competent to produce the actual linear acceleration of the center of mass C. Hence, from §83, *its component in any direction must equal the sum of the components of the acting forces in that direction.*

2. The resultant must be competent to produce the actual angular acceleration about any axis. Hence, *its moment about any axis must equal the sum of the moments of the acting forces about that axis.*

If a force satisfies these conditions it is the resultant. We shall now apply these tests to find the resultant of the forces acting on a body in some cases of importance.

101. Resultant of Two Parallel Forces in Same Direction.—Let the magnitudes of the forces be P and Q, the magnitude of their resultant being R. From (1) of §100 it is evident that the resultant must be in the same direction as the forces and equal to their algebraic sum. From (2) of §100 it follows that the resultant must be in the plane of the forces. If it were not, it would have a moment about an axis that intersects the lines of action of the forces (but not that of resultant), whereas the moments of the forces about such an axis are zero.

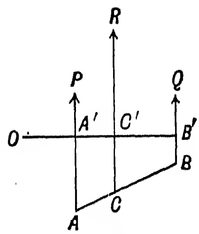


FIG. 47.

In Fig. 47 the forces are in the same direction and are applied at A and B. O is the point in which any axis perpendicular to the plane of the forces cuts the plane, $OA'B'$ is a line perpendicular to the forces, and C' is the point in which the line of action of the resultant cuts $A'B'$. Then, from the two conditions to be satisfied by the resultant,

$$R = P + Q$$
$$R \cdot OC' = P \cdot OA' + Q \cdot OB'$$

These equations enable us to find R and OC', when P, Q, OA', OB' are given.

The position of C' can be found in another way. Since the moment of the resultant about C' is zero, the moments of the forces about C' must be equal and opposite. Hence

$$P \cdot C'A' = Q \cdot C'B'$$

If C is the point in which the line of action of the resultant cuts AB, CA and CB are in the same ratio as $C'A'$ and $C'B'$. Hence

$$P \cdot CA = Q \cdot CB$$

that is, the resultant cuts AB in a point C that divides AB into parts that are *inversely as the forces.* It should be noted, for a later purpose, that this is also the rule for finding the center of mass of two particles (§78). Moreover, the position of C in AB depends only on the magnitudes of the parallel forces and is independent of their actual directions.

102. Parallel Forces in Opposite Directions.—In this case (Fig. 48) the letters have the same meanings as before. Suppose P to be greater than Q. The point C' in which the line of action of the resultant cuts $A'B'$ cannot be between A' and B', for the

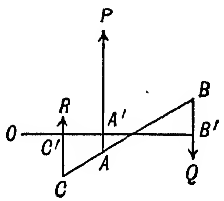


FIG. 48.

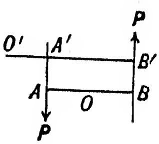


FIG. 49.

moments of the forces about C' would then be of the same sign and could not, when added, be zero. C' must be in $B'A'$ extended; it must be nearer A' than B', since P is greater than Q. Hence C is in BA produced through the point A at which the greater force acts. From the conditions of §100 we get in this case

$$R = P - Q,$$
$$R \cdot OC' = P \cdot OA' - Q \cdot OB'$$

Taking moments about C' gives

$$P \cdot C'A' = Q \cdot C'B'$$

and therefore

$$P \cdot CA = Q \cdot CB$$

If P and Q are nearly equal, the ratio $CA:CB$ must be nearly 1, which means that C must be at a great distance from A and B. Moreover, in this case R is a small force. This leads us to consider the case in which P and Q are equal.

103. Couples.—Two equal and opposite forces, not in the same line, constitute a *couple.* If we attempted to find the resultant of two such forces by the method of the last section, it would give zero force at an infinite distance, and such a force has no real existence. Hence a couple cannot be reduced to a single force.

The sum of the moments of two forces constituting a couple is the same about all axes perpendicular to the plane of the couple. For, if the axis is through O (Fig. 49), the moments of the forces are in the same direction, and their sum is $(P \cdot OA + P \cdot OB)$ or $P \cdot AB$; and, if the axis is through O', the moments are in opposite directions, and their sum is $(P \cdot O'B' - P \cdot O'A')$ which again equals $P \cdot AB$.

The distance AB between the forces of a couple is sometimes called the *arm* of a couple, and the moment of the couple about any axis perpendicular to its plane, that is $P \cdot AB$, is sometimes called the *strength* of the couple. Two couples in the same or parallel planes and of the same strength are equal in all respects and produce equal effects.

Since the algebraic sum of the forces of a couple equals zero, the couple produces no acceleration of the center of mass (§83). If the center of mass be at rest, it will remain at rest, or if it be moving in any way, it will continue moving with constant velocity. The angular velocity produced by the couple must therefore be about some axis through the center of mass.

104. Resultant of Any Number of Parallel Forces.—To find the resultant of any number of parallel forces, whether in one plane or not, we may find the resultant of two, then combine this resultant with a third, and so on. The final resultant will be either a single force or a couple or zero. At each step the resultant equals the algebraic sum of the forces added. Hence **the resultant of all the forces equals the algebraic sum of all the forces.**

The line of action of the resultant may also be found by applying the principle that **the moment of the resultant about any axis must equal the sum of the moments of the forces about that axis.** When the forces are all in one plane, to find the line of action of the resultant we only need to take moments about any axis perpendicular to the plane. When the forces are not all in one plane, it will be necessary to take moments about two rectangular axes perpendicular to the forces.

105. Center of Gravity.—Attention has been called (end of §101) to the identity of the method of finding the resultant of parallel forces in the same direction and the method of finding the center of mass of a number of particles. If, for the particles in a certain group of particles or of a body, we substitute parallel forces, all in one direction, acting at the respective positions of the particles, and proportional to the masses of the particles, the point of action of the resultant will coincide with the center of mass. It should be noticed that nothing need be said as to the common direction of the parallel forces.

The forces of gravity on the particles of a body are (very nearly) parallel forces, and they are proportional to the masses of the particles. Hence the *center of gravity* of a body, or the point of action of the resultant of the (very nearly) parallel forces of gravity, coincides with the center of mass of the body.

A very large body near the earth has a definite center of mass, but not a definite center of gravity (except in some particular cases), for the forces are not quite parallel, nor quite proportional to the masses. This is of no practical importance as regards bodies of the size found on the earth's surface; but it is of great importance in considering the effect of the attraction of the sun and moon on the motion of the earth.

106. Centrifugal Force.—In §48 we found an expression for the force required to keep a particle revolving in a circle. We may now extend this to *a body of any size or shape.* When a body of mass m rotates with constant angular velocity about any axis not through the center of mass, the center of mass moves uniformly in a circle and has therefore an acceleration v^2/r toward the center. Hence the force acting on the body (or the resultant of the forces, if there are several) must, by the principle stated in §83, equal mv^2/r, and must act in the line joining the center of mass to the center of the circle, and the body will react with an equal and opposite force. This reaction is the cause of the varying force that an unbalanced flywheel exerts on the axis.

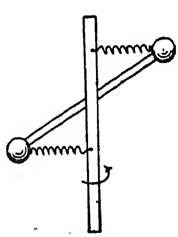


FIG. 50.

In many cases more than a single force (in addition to those required to overcome friction and sustain the weight of the body) is required to keep a body rotating about an axis. As a simple case consider a pair of equal spheres joined by a light rod and rotating

about a vertical axis through the center of the rod. (Fig. 50.) Since the center of mass has no acceleration, the forces acting on the body, if transferred to the center of mass, would have a zero resultant. Hence the forces must form a couple, and the reactions on the axis will form a couple, called a *centrifugal couple*, tending to bend the axis or make it rotate about an axis perpendicular to itself. For certain axes of rotation of a body, the centrifugal couple is zero. In the above simple illustration, this is true when the axis of rotation is in the line of the centers of the balls or at right angles thereto. These are also the positions of maximum and minimum rotational inertia of the body. A similar statement will evidently apply to a symmetrical body, such as a circular disk, that can be divided into pairs of particles like the above. Whatever the shape of a body, there are three rectangular axes through any point of a body about which it can rotate without exerting any centrifugal couple. These are the axis of maximum rotational inertia through the point, that of minimum rotational inertia, and a third perpendicular to both. These are called the *principal axes* through the point.

When a body is set spinning about a principal axis through its center of mass, it continues to spin without any tendency to "wobble," or exert a centrifugal couple. This is illustrated by the motion of a well-thrown quoit or discus, by that of a bullet from a rifled gun, and by the motion of the earth about its axis. But when the axis of initial spin is not a principal axis, irregular motion ensues, as is illustrated by a badly thrown quoit.

FORCES IN EQUILIBRIUM

107. Conditions of Equilibrium.—The forces acting on a body are in equilibrium when they cause no acceleration either linear or angular, that is, when their resultant is zero.

Given that a system of forces is in equilibrium we may conclude (from §83) that **the sum of their components in any direction equals zero,** since there is no acceleration of the center of mass, and also that **the sum of their moments about any axis equals zero,** since there is no angular acceleration about any axis.

When we equate the sum of the components of the forces in *any* direction to zero we get a relation between the forces, and it might seem that we could get an unlimited number of such relations;

but, in reality, there are only three of these independent, *e.g.*, those got by taking the sum of the components in some three directions at right angles.

Similarly, we get a relation between the forces by equating the sum of the moments about *any* axis to zero; but again there are only three of these relations independent, *e.g.*, those got by taking moments about some three rectangular axes.

Thus we can deduce, at most, *six independent relations* between forces in equilibrium, and this might have been expected from the fact that a rigid body has six degrees of freedom, at most,—three of translation and three of rotation.

We may reverse the point of view, and ask what relations and how many must forces satisfy to make it certain that they shall be in equilibrium, that is, what are the conditions essential to equilibrium. The answer is again six relations, namely, the sum of the components in each of any three rectangular directions must equal zero, and the sum of the moments about each of some three rectangular axes must equal zero.

108. Forces in a Plane.—When the lines of action of forces that are in equilibrium lie in one plane, **the sum of the components of the forces in each of any two directions at right angles in the plane equals zero.** In this case the third rectangular axis is perpendicular to the plane, and the component of each force in that direction is zero. Also, **the sum of the moments of the forces about any axis perpendicular to the plane is zero.** The other two rectangular axes are in the plane, and the moment of any one of the forces about such an axis is zero.

Hence, when forces in a plane are in equilibrium, *three independent relations among the forces can be deduced.* When we write down these relations among the forces that keep a body in equilibrium, we must include *only the forces that act on the body.* To every force that acts on the body, the body exerts an equal and opposite reaction, but these reactions are forces that *act on some other body.*

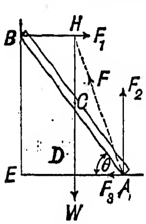


Fig. 51.

109. Example of Equilibrium of Forces in a Plane. To illustrate the above, we shall consider an example.

A uniform beam AB (length $= l$) rests, without slipping, on the ground, and leans, without friction, against a smooth wall. What is the force, F_1, exerted on the beam by the wall and the vertical force exerted on the beam by the ground,

F_2, and what is the force of friction, F_3, exerted on the beam by the ground (Fig. 51)?

Since there is no friction at B, F_1 is horizontal. The force of friction at A, that is F_3, is horizontal and toward E. Equating the sum of the horizontal forces *acting on the beam* to zero, we get

$$F_1 - F_3 = 0 \tag{1}$$

and, equating the vertical forces that *act on the beam* to zero, we get

$$F_2 - W = 0$$

A third relation may be obtained by taking moments about *any* axis perpendicular to the plane of the forces. If we choose for this purpose an axis through A, the relation will be as simple as possible, since F_2 and F_3 have zero moment about such an axis. The weight acts at the center C of the beam, and the distance of its line of action from A is $(l/2)\cos\theta$. Also the distance, BE, of the line of action of F_1 from A equals $l\sin\theta$. Hence

$$W\frac{l}{2}\cos\theta - F_1 l\sin\theta = 0 \tag{3}$$

From these three equations we get

$$F_1 = F_3 = \tfrac{1}{2}W\cot\theta$$
$$F_2 = W$$

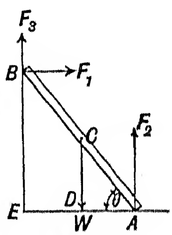


FIG. 52.

110. Another Example of Equilibrium.—A uniform rod hangs from a wall by a hinge and rests on a smooth floor (Fig. 52). In this case the force at A must be vertical, since there is no horizontal force of friction at A. Let the force on the beam at B consist of a horizontal part F_1 and a vertical part F_3. Equating to zero the sum of the vertical forces, the sum of the horizontal forces, and the sum of the moments about B, we get

$$F_1 = 0 \qquad F_2 + F_3 - W = 0$$

$$W\frac{l}{2}\cos\theta - F_2 l\cos\theta = 0$$

Hence

$$F_1 = 0 \qquad F_2 = \tfrac{1}{2}W \qquad F_3 = \tfrac{1}{2}W$$

Since F_1 is zero, the rod does not press against the wall. This result, which seems at first improbable, may be verified by allowing A to rest on a board in a tank of water and hanging B by a cord: the cord will be found to be vertical when tested by comparison with a plumb line.

111. Special Cases of Equilibrium.—1. *When two forces are in equilibrium, they must be equal and opposite and in the same line.* If not equal and opposite, they would produce translation; and if not in the same line, they would produce rotation.

For example, a body suspended by a cord must rest so that its center of gravity is vertically below the point of support. This supplies an experimental method of finding the center of gravity of a disk of any shape. It is only necessary to support it in succession at two points on its rim and find the intersection of the lines of support.

2. *When three forces are in equilibrium they must all lie in one plane.* For the sum of the moments of all three about any axis is zero. About any axis that intersects the lines of action of two of the forces the moments of these two forces are zero. Hence any such axis must also intersect the line of action of the third force (unless it be parallel to it), and this cannot be so unless all the forces lie in one plane.

3. *Three forces in equilibrium must either be parallel or pass through a single point.* If they are parallel, one is equal and opposite to the resultant of the other two. If they are not parallel, two of them intersect, and their moments about any axis through the point of intersection are zero. Hence the third must pass through the point of intersection of any two.

As a case of three parallel forces in equilibrium, consider the example in §110. The resultant of F_2 and F_3 must be equal and opposite to and in the same line as W, which acts at the middle of AB. Hence F_2 and F_3 are equal.

As a case of three non-parallel forces in equilibrium, consider the example in §109. Let the resultant of F_2 and F_3 be F. Then F, F_1 and W are three forces in equilibrium. Hence F must pass through the intersection of F_1 and W. Thus the direction of F is readily found graphically. We may also find graphically the magnitudes of F_1 and F. Since DA and BH are equal, $HBDA$ is a parallelogram. Hence F, F_1 and W are proportional to AH, BH and HD.

112. Stable, Unstable and Neutral Equilibrium.—A body is in equilibrium when it is either at rest or moving uniformly, that is, without acceleration, linear or angular. The resultant of the forces acting on such a body is zero.

When a body in equilibrium is at rest, the equilibrium is described as *static*. Of this kind of equilibrium there are three forms, stable, unstable and neutral. A body at rest is in *stable equilibrium* when, on being slightly displaced, it tends to return to its equilibrium position. This is illustrated by a chemical balance, a pendulum or picture hanging by a cord, a book on a table, and, in fact, by most stationary objects. A body at rest is in *unstable equilibrium* when, on being slightly displaced, it tends to move further from its equilibrium position. An egg on end and a board balanced on one corner would be in unstable equilibrium. A body at rest is in *neutral equilibrium* when, on being slightly displaced, it has

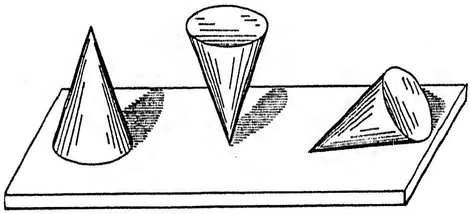


FIG. 53.—Stable, unstable and neutral equilibrium.

no tendency either to move further away or to return; for example, a sphere or cylinder on a horizontal table and any body mounted on an axis through its center of gravity.

A body in a position of stable equilibrium oscillates about that position when displaced and released, though the oscillation may be quickly destroyed by friction or other forces. When too far displaced, such a body may come to a position of unstable equilibrium and not return; a table or chair tilted too far comes to a position of unstable equilibrium. The extent to which any such body may be displaced and yet return is a measure of the degree of stability of the equilibrium.

113. Energy Test of Static Equilibrium.—When a body at rest is in stable equilibrium, a disturbance will increase its potential energy. This is evident in the case of a pendulum at rest, for a disturbance raises its center of gravity; work is done against gravity when the body is displaced, and this work produces potential energy.

Thus **a position of stable equilibrium is a position in which the potential energy is a minimum.** This statement holds true whatever the force against which work is done; the fact that the body, when displaced, tends to return shows that there are conservative forces opposing the motion. Work done against these results in an increase of potential energy.

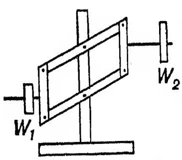


FIG. 54.—Neutral equilibrium.

A position of unstable equilibrium is a position in which the potential energy is a maximum, as is illustrated by a spheroid on end or a board balanced on a corner; a disturbance lowers the center of gravity. The statement is true whatever the forces in action; for the fact that the body, when disturbed, moves farther away from its position of equilibrium, and thus gains kinetic energy, shows that its potential energy diminishes.

When the equilibrium is neutral, a displacement produces no change of potential energy; when a sphere rolls on a horizontal table its center neither rises nor falls. An interesting illustration is afforded by the apparatus sketched in Fig. 54. It will remain at rest whatever the positions of the equal weights, which are adjustable along the horizontal rods, for the total potential energy is the same in all positions. This is the principle of a common form of balance. The weights or the body weighed may be placed on any part of the pan.

A body, such as a fly-wheel or a railway car, in a steady state of motion is in *kinetic equilibrium* since the resultant of the forces acting on it is zero.

The principle that, for stable equilibrium, the potential energy is a minimum is extensively illustrated in nature; the potential energy may be partly or wholly other than mechanical energy, in forms dealt with in other parts of Physics. Changes are continually taking place in nature, and bodies, when disturbed, settle into states of stable equilibrium, that is, of minimum potential energy.

PERIODIC MOTIONS

114. A periodic motion is one that is repeated in successive equal intervals of time. The time required for each such repetition is called the *period* of the motion. Thus, the moon revolves around the earth with a periodic motion, the period of which is a lunar month, and the earth revolves about the sun in a period of a year. The end of a hand of a clock has a periodic motion about

the center of the face. A point on a vibrating violin string or piano wire has a periodic motion.

115. Uniform Circular Motion.—When a point P revolves with constant speed in a circle of center O, the position of P at any moment may be assigned by giving the angle that OP makes with some fixed radius, such as OA. This angle is called the *phase* of P's motion.

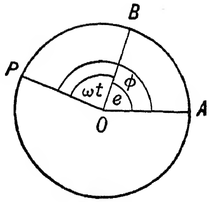


FIG. 55.

If the period of the motion is T, the angle through which OP revolves in unit time is the angular velocity ω and equals $2\pi/T$. Let us suppose that, at the moment from which we begin reckoning time, P is at some position B, and let its phase at that moment, that is, the angle BOA, be e. After time t, P will have revolved through an angle ωt or $(2\pi/T)t$ and the phase at time t will be $[(2\pi/T)t + e]$.

116. Simple Harmonic Motion.—This is the most important form of periodic motion and is illustrated by the vibration of a simple pendulum swinging in a small arc, of a weight hung from an elastic cord or spring and moving vertically, of a point on the prong of a tuning fork, and many other cases.

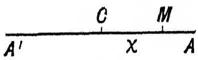


FIG. 56.

Simple harmonic motion is a **linear vibration, the motion being such that the vibrating point has an acceleration that is toward the center of its path and proportional to its distance from the center.**

Let $A'A$ be the path of vibration of a point M that has a simple harmonic motion, and let C be the center of $A'A$. Denote the distance of M from C at any time by x, and let values of x be considered as positive when M lies between C and A and negative when M lies between C and A'. When x is positive, the acceleration, a, of M is toward C and is, therefore, in the negative direction, and when x is negative, a, being still toward C, is positive. Hence, if we denote the constant of proportionality of the magnitude of a to x by c, by the above definition of simple harmonic motion

$$a = -cx$$

The distance, x, of the vibrating point from the center of motion at any moment is called the **displacement** *of the point at that moment.*

One-half of the length of the path of vibration is called the **amplitude** *of the simple harmonic motion.* We shall denote it by r. It is equal to the magnitude of the greatest displacement (CA or CA').

The time required for a complete vibration (that is, from A to A' and back to A) is the **period** of the simple harmonic motion.

117. The Force Acting on a Body that has Simple Harmonic Motion.—A body that has a simple harmonic motion has a varying acceleration that is always directed toward the center. To produce this acceleration, a varying force, also directed toward the center, must act on the body. Denote the force by F, and let m be the mass of the body. From the Second Law of Motion and the definition of simple harmonic motion, we get

$$\begin{aligned} F &= ma \\ &= -mcx \end{aligned}$$

Since m and c are constants for a given body and a given simple harmonic motion, **the force required is always opposite to and proportional to the displacement.**

The force required to stretch or compress a spiral spring, one end of which is fixed, is proportional to the displacement of the free end from its unstrained position, and the reaction exerted by the spring is opposite to and proportional to the displacement (§60). Hence a body attached to such a spring and allowed to vibrate under the action of the spring has simple harmonic motion. The same law of force holds for a flat spring when bent, and, in fact, for any elastic body when distorted. Hence all elastic vibrations are simple harmonic motions or compounded of such motions, and the same is true of the vibrations that constitute sound and light.

118. Relation Between Simple Harmonic and Circular Motions. Simple harmonic motion has been defined as a vibration in a line according to the law $a = -cx$. Now *the projection of a uniform circular motion on a diameter of the circle has exactly the same character.* For, on $A'A$ as diameter draw a circle (Fig. 57), and let P revolve with constant angular velocity, ω, in the circle. If M is the projection of P on $A'A$, M vibrates once along $A'A$ in each revolution of P. Since the motion of M is that part of the motion of P that is in the direction of $A'A$, the acceleration, a, of M is the component of the acceleration of P in that direction. The acceleration of P

is $\omega^2 r$ in the direction PC, or $-\omega^2 r$ in the direction of CP. Hence

$$a = -\omega^2 r \cos\theta = -\omega^2 x$$

Since ω is constant throughout the motion, the projection is a simple harmonic motion, in which $c = \omega^2$. Hence **any simple harmonic motion may be regarded as a projection of a uniform motion in a circle.** The circle is called the *circle of reference* of the simple harmonic motion.

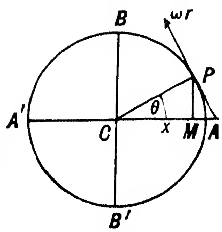


FIG. 57.

This relation between circular motion and simple harmonic motion affords a means of deducing some of the properties of simple harmonic motion, without the use of advanced mathematics.

119. Period of a Simple Harmonic Motion.—Consider a simple harmonic motion as a projection of a uniform circular motion. The periods of the two motions must be the same. Denote it by T. From §118

$$a = -\omega^2 x = -\left(\frac{2\pi}{T}\right)^2 x$$

Hence

$$T = 2\pi\sqrt{-\frac{x}{a}} = 2\pi\sqrt{\frac{1}{c}}$$

Since x and a are always of opposite signs, the quantity under the radical is always numerically positive.

120. Displacement in a Simple Harmonic Motion.—From the same relation between circular motion and simple harmonic motion, we can also deduce an expression for the displacement at any moment in the simple harmonic motion. Let P be the point whose motion projects into that of the vibrating point M (Fig. 58). The angle PCA, or the phase of P's motion (§115), equals $[(2\pi/T)t + e]$. Hence for CM or the displacement, x, in the simple harmonic motion, we have

$$x = r\cos\left(\frac{2\pi}{T}t + e\right) \qquad (1)$$

While we have deduced this expression from the circular motion, it must now be regarded as an expression for the simple

harmonic motion of amplitude r and period T. $[(2\pi/T)t + e]$ is called the *phase* of the simple harmonic motion at time t, e being the phase of the simple harmonic motion at zero time.

For two particular values of e the expression for x becomes simpler. If e is zero, which, as we see from the circular motion, means that at zero time M is at A, the expression for x is

$$x = r \cos \frac{2\pi}{T}t \tag{2}$$

If $e = -(\pi/2)$, at zero time P is at B, and M is therefore at C and moving in the positive direction. Substituting this value of e in the above general expression for x, we get

$$x = r \sin \frac{2\pi}{T}t \tag{3}$$

We might have started with (2) or (3) as *defining* simple harmonic motion. By differentiation it gives $a = d^2x/dt^2 = -cx$ where $c = (2\pi/T)^2$. Hence the two definitions are equivalent.

121. Velocity in a Simple Harmonic Motion.—The velocity in a simple harmonic motion is obviously a *variable velocity*. When the vibrating point is at A or A', (Fig. 58) it is, for a moment, at rest, and its velocity, v, is zero. In the motion from A' to A the velocity is positive, while from A to A' it is negative. When the point is at C, the magnitude of its velocity is a maximum, and, since it is then the full velocity of P, it is equal to $2\pi r/T$. Hence v is a maximum where a is a minimum, and v is a minimum where a is a maximum.

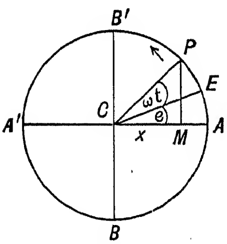


Fig. 58.

An exact expression for v in terms of x as a variable can readily be obtained by taking the component, along CA, of P's velocity, $2\pi r/T$, along the tangent to the circle, but it is not frequently needed and may be left as an exercise.

It can also be found by differentiating the expression for x in §120.

122. Simple Pendulum.—A simple pendulum consists of a small heavy body, called the bob (usually spherical), suspended by a practically inextensible cord, the mass of which is so small as to be negligible compared with the bob. As the pendulum swings through a small angle, the bob vibrates through a small circular arc, which is very nearly a straight line.

The force of gravity, mg, on the bob of the pendulum acts vertically, and it may be resolved into a component along the tangent and a component along the radius. The latter component produces a tension in the cord, which does not affect the motion, while the former component produces an acceleration along the tangent. When the cord is at an inclination θ to the vertical, the component along the tangent is $mg \cos [(\pi/2) - \theta]$, or $mg \sin \theta$. Since the pendulum is supposed to vibrate through a very small angle, $\sin \theta$ may be replaced by θ; in fact, for values of θ less than 2°, $\sin \theta$ and θ do not differ by more than one part in 10,000. If the distance of the bob from its lowest point, measured along the arc, be denoted by x, and the length of the pendulum by l, $\theta = x/l$ radians. Hence the force along the tangent is $mg(x/l)$. This force is in the negative direction when x is positive. Hence, denoting the acceleration along the tangent by a, we have by the Second Law of Motion

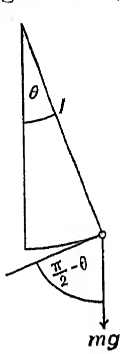


FIG. 59.—Simple pendulum.

$$-mg\frac{x}{l} = ma$$

and

$$a = -\frac{g}{l}x$$

Since the multiplier of x is a constant throughout the motion, the acceleration is opposite to and proportional to the displacement. Hence the motion is simple harmonic motion, and, if T be the period, or time of vibration, of the pendulum, by §119

$$T = 2\pi\sqrt{-\frac{x}{a}}$$
$$= 2\pi\sqrt{\frac{l}{g}}$$

In this equation T means the time of a *complete* vibration, to and fro, but the clockmakers' term "second's pendulum," still frequently used, denotes a pendulum that makes a complete vibration in 2 seconds.

123. Angular Harmonic Motion.—A body attached to an axis may vibrate backward and forward through an angle, as in the case of a balance wheel of a watch or of any heavy body hung on

a peg. When the angular acceleration, α, is always opposite to and proportional to the angular displacement, θ, the motion is called *angular harmonic motion.* Hence the general formula for such motion is

$$\alpha = -c \cdot \theta$$

c being a constant.

Figure 60 is supposed to be drawn in a plane through the body perpendicular to the axis O, and M is a point in the vibrating body. A line OM in the body will vibrate backward and forward through an angle. The point M will vibrate in an arc of a circle of radius OM or r. When the angular displacement of OM from its mean position, OC, is θ, the displacement, x, of M from C is $r\theta$ and the linear acceleration, a, of M is $r\alpha$ (§74). Substituting these values of θ and α in the above formula and cancelling r, we get

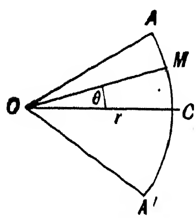


FIG. 60.

$$a = -c \cdot x$$

Thus the motion of M is simple harmonic motion in all respects except that it is along an arc (which may be long or short) instead of along a straight line. We might suppose the arc straightened out without any other change in the nature of the motion of M. Hence, if T be the period of M's motion, which, of course, is the same as the period of the angular harmonic motion,

$$T = 2\pi\sqrt{-\frac{x}{a}}$$

$$= 2\pi\sqrt{-\frac{\theta}{\alpha}}$$

This expression for the calculation of the period of an angular harmonic motion is similar to that for the calculation of the period of a simple harmonic motion (§119).

As examples of angular harmonic motion we shall consider the torsion pendulum and the compound pendulum.

124. The Torsion Pendulum.—A torsion pendulum consists of a vertical wire carrying a body at one end and clamped at the other end (Fig. 61). When the body is turned around the wire as axis and released, it performs angular vibrations; the twisted wire

begins to untwist and thus starts the motion, which persists after the wire has untwisted, owing to the kinetic energy acquired by the body.

To twist the wire requires the application of a couple. The twist, θ, produced by a certain couple of moment L, is proportional to L and it also depends on the material and dimensions of the wire. Hence $L = K\theta$, where K is a constant for this particular wire, called the *constant of torsion* of the wire. The couple exerted by the twisted wire is equal and opposite to that required to produce the twist. Hence the couple exerted by the wire on the body is $-K\theta$ when the displacement is θ. This couple gives the body an angular acceleration, and if we denote this by α and the rotational inertia of the body by I,

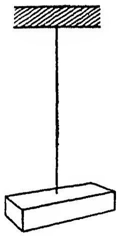


Fig. 61.—Torsional pendulum.

$$-K\theta = I\alpha$$

and

$$\alpha = -\frac{K}{I}\theta$$

Since the multiplier of θ is a constant, the motion agrees with the definition of angular harmonic motion, and, if T is the period of vibration,

$$T = 2\pi\sqrt{-\frac{\theta}{\alpha}}$$

$$= 2\pi\sqrt{\frac{I}{K}}$$

It should be noticed that we have not assumed the angle of vibration to be small, as in the case of the ordinary pendulum; in the torsion pendulum the restoring couple is proportional to the angular displacement, even when the latter is large (provided it is not so large as to permanently strain the wire).

By means of the torsion pendulum the rotational inertia of an irregular body can be compared with that of a body of known rotational inertia. The two are, by the above formula for T, proportional to the squares of the corresponding times of vibration, when the bodies are, in turn, attached to the same wire and set into angular vibration.

125. The Compound Pendulum.—A body of any shape, suspended by a horizontal axis and vibrating under gravity through a small angle, constitutes a compound pendulum. Fig. 62 represents a vertical section through the center of gravity C and perpendicular to the axis of suspension, which it intersects in S. The point S is called the center of suspension. Denote SC by h. When SC is inclined at an angle θ to the vertical, the force of gravity, mg, which acts at C, has a moment about the axis through S which is equal to $mgh \sin \theta$ and is negative when θ is positive. This is the only moment of force about the axis. Hence, if I is the rotational inertia of the body about the axis through S,

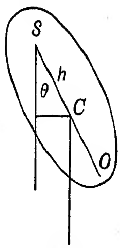


FIG. 62.

$$-mgh \sin \theta = I\alpha$$

If the angle θ is always small, we may, as in the case of the simple pendulum, replace $\sin \theta$ by θ and thus get

$$\alpha = -\frac{mgh}{I}\theta$$

This satisfies the condition for angular harmonic motion, and the period of vibration is

$$T = 2\pi\sqrt{-\frac{\theta}{\alpha}}$$

$$= 2\pi\sqrt{\frac{I}{mgh}}$$

In this equation $I = I_0 + mh^2$ where I_0 is the rotational inertia of the body about an axis through C parallel to the axis through S.

126. Center of Oscillation. Reversible Pendulum.—Let the radius of gyration about an axis through C, parallel to the axis of suspension, be k. Then the rotational inertia about the axis through C is mk^2, and about the parallel axis through S it is $mk^2 + mh^2$ (§§89, 94), which is, therefore, the value of I. Hence

$$T = 2\pi\sqrt{\frac{k^2 + h^2}{gh}}$$

If this be compared with the formula for a simple pendulum, it is seen that, if l be the length of a simple pendulum that has the same period of vibration as the compound pendulum,

$$l = \frac{k^2 + h^2}{h} = \frac{k^2}{h} + h$$

Hence $$(l - h)h = k^2$$

The length l is evidently greater than h. Hence, if we measure along SC a length equal to l, we shall arrive at a point O in SC extended. The point O, which is always on the opposite side of C from S, is the point at which the whole mass of the body might be supposed concentrated, without any alteration of the period of vibration. O is called the **center of oscillation** corresponding to the center of suspension S. Since $CO = (l - h)$ and $CS = h$, we have, as the relation between any center of oscillation and the corresponding center of suspension,

$$CS \cdot CO = k^2$$

If the pendulum be now inverted and set to vibrate about an axis through O, parallel to the former axis, the new center of oscillation, O', will lie in OC produced and must satisfy the relation

$$CO \cdot CO' = k^2$$

A comparison of these two equations shows that O' must coincide with S Hence *the center of suspension and the center of oscillation are interchangeable, and the distance between them is the length of the equivalent simple pendulum.* This is the principle of Kater's *reversible* pendulum. Because of its rigidity it is used instead of a simple pendulum in geodetic surveys.

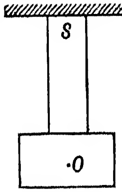


Fig. 63.

127. Energy Changes.—The resultant force of gravity acts at C (Fig. 62). Hence the potential energy of the pendulum, in any position, is the same as if its mass were concentrated at C. But the pendulum does not swing as if it were concentrated at C, because its kinetic energy is that of its mass supposed concentrated at C plus its kinetic energy of rotation about C (§95).

As the pendulum falls toward the vertical, the lost potential energy goes partly into energy of rotation about C; hence it does not swing as rapidly as if it were concentrated at C, that is, as if it were a simple pendulum of length SC. A parallel case that brings out the distinction is illustrated by a block suspended by two cords as in Fig. 63. Swinging perpendicularly to the plane of the figure it is a physical pendulum of length SO, the block having energy of rotation. Swinging parallel to the plane of the figure, it is a simple pendulum of length equal to the length of the cords; the block in this case has no rotation. A similar explanation applies to the motion of the pans of a balance. They do not rotate with the beam but move vertically; hence they affect the motion as if concentrated on the supporting knife-edges.

Fig. 64. Center of percussion.

128. Center of Percussion.—There is another important relation between a center of suspension S and the corresponding center of oscillation O (Fig. 64). A blow at O transverse to SO will start the body rotating about S, without any jar on the support at S. Hence O is also called the *center of percussion* of the body when suspended at S. The center of percussion may be found by holding the body at S and striking it transversely.

When the blow is through the center of percussion, there is no jar on the hand. When a baseball hits the bat at the center of percussion, there is no jar on the hands.

129. Gyroscopic Motion.—A gyroscope, in its simplest form (Fig. 65), is a wheel on a horizontal axle, which is supported on a pivot (a bicycle wheel suspended by a vertical cord attached to a short extension of the axle will serve). When the wheel is set in rotation and the axle then released, the axle, instead of tilting in a vertical plane, as it would if the wheel were at rest, revolves in a horizontal plane at a rate that depends on the velocity of rotation of the wheel about the axle. This motion is called **precession.** (Vertical oscillations or *nutations* of the free end of the axle may also accompany the precession.) The weight of the wheel, acting

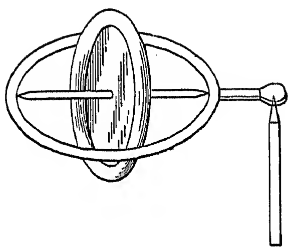


Fig. 65.—A gyroscope.

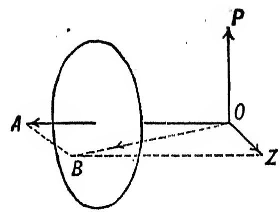


Fig. 66.

at the center of the wheel, has a moment about an axis through the pivot at right angles to the axis of the wheel. If this moment of force be increased by hanging a weight on the frame, the rate of precession will be greater. If the wheel be supported at its center of gravity, there will be no moment of force exerted on it by gravity, and no precession will take place, that is the axis will continue to point in a *fixed direction in space.* Such a suspension is obtained by using double gimbal rings for the support of the axis. This principle was used by Foucault to demonstrate the rotation of the earth.

The motion of a gyroscope (Fig. 65) is analogous to the revolution of a particle in a circle, under the action of a force directed toward the center. The latter requires a force *perpendicular to the direction of motion,* while precession requires a moment of force about an axis *perpendicular to the axis of rotation.*

The principal properties of a body acting as a gyroscope may be considered to be the following:

(1) When there is no moment of force acting on the gyroscope, the axis of rotation maintains a fixed direction in space; (2) when some external agent exerts a moment of force at right angles to the axis of rotation, precession takes place and (3) the gyroscope then exerts an equal and opposite reaction moment on the external agent; (4) a moment of force about an axis parallel to the axis of precession will cause a tilt of the gyroscope.

130. Moment of Force and Rate of Precession.—Let I and ω be the rotational inertia and the angular velocity of a gyroscope about its axis. Let L be the moment of force that causes the precession and ω' the angular velocity of precession. To find the relation between these quantities let OA (Fig. 66) represent the angular momentum, $I\omega$. After a short time, t, OA will have turned through a small angle ϕ to the position OB, where $\phi = \omega' t$. Hence angular momentum, represented by OZ or AB, must have been added, and this must equal Lt (§91).

$$\therefore \frac{Lt}{I\omega} = \frac{AB}{AO} = \omega' t.$$

Hence $$L = I\omega\omega'$$

Thus ω' is proportional to L and inversely proportional to $I\omega$.

131. Other Examples of Precession.—The curvature of the path of a coin rolled with a tilt along a table is due to the precession of its axis, caused by the moment of its weight about the point of contact with the table. The motion of a top is a precession due to gravity and to friction at the support.

Any large body, such as a dynamo armature, in rotation aboard a vessel that is rolling, pitching, or turning, has a precessional motion, and the bearings must supply the necessary moment of force and experience an equal and opposite reaction. Similar statements apply to the propeller of an airplane or airship.

The flywheel on the crankshaft of an automobile acts as a gyroscope. When the car is travelling on a curve, the bearings of the shaft compel a precession of the flywheel, and the shaft tends to tilt. If the rotation of the wheel is counterclockwise as seen from the rear, the front of the car tends to rise when the car is turning to the right.

The earth is not quite spherical but bulges at the equator. On one side the protuberance is closer to the moon than the center of the earth is and on the other side it is farther away. The result of this is a moment of force that (with a similar but smaller moment exerted by the sun) causes a precession of the earth's

axis, which travels on the surface of a cone and completes a revolution once in about 25,800 years.

The gyroscope has been applied to steering torpedoes, to preventing the rolling of ships, to balancing trains on a single rail, and in the construction of a non-magnetic mariner's compass. The principles used are essentially those enumerated in §129, but the mechanical details are usually very complex.

FRICTION

132. Static Friction.—When two solids are in contact, there is a resistance, caused by the surfaces, to the sliding of one on the other (Fig. 67). This resistance is called *friction.* When a force parallel to the surfaces of contact is applied to one of the bodies, and the force is less than a certain amount, motion will not take place, the resistance being equal to the force. When the force is increased to a certain value, the resistance will fail to increase, and sliding will take place. This maximum resistance is called the *maximum static friction.* It depends on the nature of the surfaces and also on the force each body exerts on the other normally to the surface of contact. This normal force we shall, for brevity, call the *thrust* across the surface or simply the thrust.

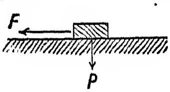


Fig. 67.

With a given pair of surfaces in contact and with a force tending to produce sliding motion in a certain direction (to take account of the influence of grain), *the maximum static friction is found to be* (within certain wide limits) *proportional to the thrust.* Denoting the coefficient of proportionality by μ, the maximum static friction by F, and the thrust by P, we have

$$F = \mu P$$

The constant μ is called the *coefficient of static friction,* and may be defined as *the ratio of the maximum static friction between two surfaces to the thrust between them.*

If one of the two bodies rests on the other, and if the surfaces of contact are plane and horizontal, the thrust is equal to the weight of the upper. The maximum static friction is equal to the force applied horizontally to the upper that will just produce motion. If additional weights be placed on the upper body, the thrust will be increased, and the friction will be increased in the same proportion. If the upper body be redistributed in any way, for instance, if it

be cut in two and one part placed on the other, the total force of friction will not change; for, while the area of contact will be diminished, the thrust on each unit of area will be increased in the same proportion. Thus *the total frictional resistance is independent of the area of contact*, and, for two given surfaces, it depends only on the thrust between them, as is implied in the equation $F = \mu P$.

133. Effect of State of Surface.—The coefficient of static friction between two surfaces depends on the materials and a variety of circumstances. The rougher the surfaces, that is the greater the inequalities in each, the larger is μ. If the surfaces are not clean, parts of the surfaces are replaced by surfaces of the foreign substance, and μ is necessarily different. The longer two surfaces are in contact, the greater is the maximum static friction: this is especially true of soft or fibrous surfaces. When the materials are of grained structure, the friction is greater across the grain than along it. Friction is usually due to interlocking of the projections on one surface with those on the other surface. When slipping takes place some projecting pieces may be broken off, or abraded, as it is called. With prolonged contact between two surfaces, small readjustments of the surface particles take place, so that the fit becomes closer, and the resistance to motion greater. It has even been found that when one surface is pushed a *very* small distance, it will, when released, spring back, thus showing that there is some elastic bending of surface projections. In general, friction between two surfaces of the same material is greater than between surfaces of different material, since the former allows more uniform interlocking. There is therefore an advantage in using brass bearings for steel shafts, to diminish friction, and covering with leather the face of a pulley, used with leather belting, increases friction, and helps to prevent slip. But interlocking is not the only cause of friction. The friction between extremely smooth surfaces, such as polished glass, in very close contact seems to be due mostly to molecular adhesion.

134. Uses of Friction.—Friction is utilized in the transmission of energy by machine belting. Usually some slipping takes place, for the belt stretches somewhat while in contact with the pulley. Brakes for automobiles and railway cars depend on friction. Friction between the driving wheels of a locomotive and the rails prevents slipping: without it the locomotive would be helpless, and where it is not sufficient the track is sanded. To hold a rope fast it is sometimes wrapped around a post. The friction on each part of the rope diminishes the tension transmitted to the next part. It is found that after one turn the tension is diminished to about $\frac{1}{9}$, after two turns to $\frac{1}{9}$ of $\frac{1}{9}$ and so on. At this rate, after five turns a pull of one pound weight on the free end would counteract a force of 4 tons at the other end.

135. Slip on an Incline.—When a body rests on an inclined plane the tilt of which is gradually increased, there is some angle, i, at which slipping begins.

The weight of the body is mg and acts vertically. It may be resolved into a component $mg \sin i$ down the plane and a component

$mg \cos i$ perpendicular to the plane. The latter component causes a thrust between the surfaces, while the former is the force parallel to the surface which produces motion. Hence, from the definition of the coefficient of static friction,

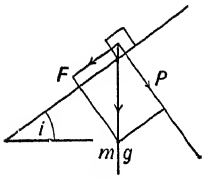


FIG. 68.

$$\mu = \frac{F}{P} = \frac{mg \sin i}{mg \cos i} = \tan i$$

Thus *the coefficient of static friction is equal to the tangent of the angle of slip.* (The angle i is also sometimes called the "angle of repose.") This relation provides a simple method of measuring μ.

136. Kinetic Friction.—To keep one body sliding on another at a constant speed, a certain force, F, parallel to the surface of contact, is required. Through a considerable range of speed this force is practically constant. The opposing resistance offered by the surfaces is called *kinetic friction.* It is found to be, for a certain pair of surfaces moving, relatively to each other, in a definite direction, proportional to the thrust, P, between the surfaces. Denoting the coefficient of proportionality by μ', we have

$$F = \mu' P$$

The constant μ' is called the *coefficient of kinetic friction.* It may be defined as *the ratio of the kinetic friction between two surfaces to the thrust between them.*

As in the case of static friction, for a given thrust between two surfaces the kinetic friction is independent of the area of contact.

The kinetic friction between two surfaces is, in general, less than the maximum static friction. The reason probably is that time is not allowed for the surface to settle into as close contact as if they were at rest. Moreover, kinetic friction is not quite independent of velocity. When the velocity is decreased until it is very small (how small depends upon the particular surfaces), the friction increases, and it continues to increase as the velocity diminishes toward zero, and at a sufficiently small velocity the kinetic friction does not differ appreciably from the maximum static friction. At very great velocities, the friction is generally less than at moderate velocities.

A friction dynamometer is a machine for measuring the power of an engine. The engine drives a wheel over which a belt hangs under known tension. From the tension of the belt and the number of revolutions made by the wheel the work done is calculated.

When a lubricant is used between two surfaces, there is no longer friction of solid on solid, and the laws of kinetic friction no longer hold: the coefficient of friction depends on both thrust and velocity and the action is very complex. The friction of a skate on ice is probably greatly diminished by the momentary liquefaction of the ice immediately under the skate, due to the great force exerted by the latter on a small area (see §301).

137. Sliding on an Inclined Plane.—A body sliding down an inclined plane (Fig. 21) is urged downward by the component of its weight along the plane and retarded by friction. If the inclination of the plane to the horizontal is i, the component of gravity along the plane is $mg \sin i$. The thrust on the plane is $mg \cos i$; hence the force of friction is $\mu' mg \cos i$. If the component of gravity down the plane exceeds the force of friction, the body will slide with an acceleration a. Hence, taking the direction down along the plane as the positive direction, we have, by Newton's Second Law,

$$ma = mg \sin i - \mu' mg \cos i$$

This suggests a method of finding μ' by measuring a and i.

138. Rolling Friction.—The term friction is also applied to the resistance experienced by a wheel in rolling on a surface without any slipping. The cause of the resistance is, in this case, entirely different. This is seen by considering the rolling of a heavy wheel on a soft substance, such as india rubber. If the wheel were at rest, it would sink into the rubber, raising a small mound on each side of the contact. When the wheel is moving forward, the mound is chiefly on the forward side, at A (Fig. 69). The thrust, P, of the rubber on the wheel at A is inclined to the vertical, in some such direction as AP. The point about which the wheel is momentarily rotating (§73) is not C but B in the figure, and the moment of P about B is necessarily opposed to that of F.

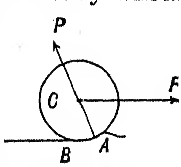


Fig. 69.—Resistance to rolling.

It follows from this explanation that the resistance to the motion is greater the softer the surface, greater the greater the thrust of

the wheel on the surface, and less the larger the wheel, since a larger wheel will distribute the thrust over a larger surface, and will not sink so deeply. When the surface on which the wheel rolls is hard, there is very little deformation, and the resistance to the motion is less. Thus, the resistance to the rolling of iron on india rubber is about ten times greater than the rolling of iron on iron. With a lignum vitæ cylinder of 16-in. diameter loaded with 1000 lbs. the rolling friction has been found to be about 3 per cent of the sliding friction when the wheel was not allowed to rotate. Because of this difference, rolling is, when possible, preferred to sliding. Thus rollers beneath a heavy body and the balls in a ball-bearing greatly diminish frictional resistance.

A pneumatic tire on an automobile flattens out in contact with the ground and does not sink in, so that it gives the wheel the advantage of a much larger wheel. But work has to be done in rapidly deforming the rubber, which recovers more slowly, and this work is at the expense of the energy of motion of the car. On a perfectly smooth, plane, hard road a pneumatic tire would be a disadvantage. On a soft rough road it is a great advantage. For a hard smooth road the tire should be pumped "hard"; for a soft road it should be "soft."

SIMPLE MACHINES

139. Machines.—A machine is a contrivance for applying energy to do work in the way most suitable for a certain purpose. The machine does not create energy; no machine can do that. To do work, it must receive energy from some store of energy, and the greatest amount of work it can do cannot exceed the energy it receives.

Different machines receive energy in different forms, some in the form of mechanical (kinetic and potential) energy, some in the form of heat energy, some in the form of chemical energy, and so on. We shall consider here only machines that employ mechanical energy and do work against mechanical forces.

In certain very elementary machines, the so-called *simple machines*, the agent that supplies the energy exerts but a single force, and the machine, at least as regards the useful work that it performs, is opposed by a single resisting force. The former may be called the *applied force* and the latter the *resistance*.

Every machine in action encounters a certain amount of frictional resistance, and the work done against it is not usually useful work. In many cases this is very small, and, in treating (to a first approximation) of the simple machines, it is customary to neglect it.

140. Mechanical Advantage.—The work done by the applied force, P, is measured by the product of P and the distance, p, through which P acts. The work done against the resistance is measured by the product of the resistance, Q, and the distance q, through which it is overcome. In a simple machine (where friction may be neglected) these must be equal. Hence

$$\frac{Q}{P} = \frac{p}{q}$$

Hence p is greater than q in the proportion in which Q is greater than P. This principle was first stated by Stevinus (1548–1620).

The ratio Q/P or the equal ratio p/q for such a machine is called the *mechanical advantage* of the machine. When friction cannot be neglected, these ratios are not equal, and, if p/q, that is, the speed ratio, is taken as the measure of the mechanical advantage, Q/P is less than the mechanical advantage.

141. Efficiency.—By the *efficiency* of a machine is meant the ratio of the useful work, or work of the kind desired, to the energy received. For a simple machine without friction this would be unity. When there is friction, the efficiency may have any value less than unity.

142. Levers.—A lever is a bar supported at a point called the *fulcrum*, F; a force P applied to the bar at a point A will overcome a resistance Q acting at another point B. We shall suppose that P and Q act at right angles to the bar and to the axis of rotation at the fulcrum.

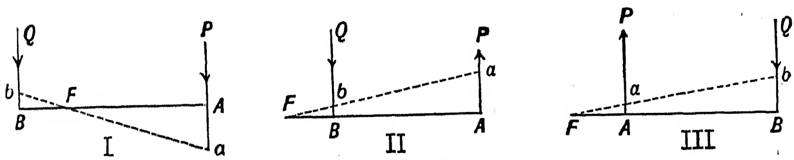


FIG. 70.—Three classes of levers.

To find the relation between P and Q, suppose the bar to turn through a very small angle, so that A moves through a distance Aa and B through a distance Bb. The work done by P is $P \cdot Aa$,

and the work done against Q is $Q \cdot Bb$. The conservation of energy requires that these should be equal. Hence

$$\frac{Q}{P} = \frac{Aa}{Bb} = \frac{AF}{BF}$$

This relation may also be found by considering the parallel forces acting on the bar or by taking moments about the fulcrum.

Levers are usually divided into three classes, represented in Figure 70. In levers of the *first class* the force, P, and the resistance, Q, are on opposite sides of the fulcrum, and the resistance may be greater or less than the applied force. To this class belong the crow-bar, forceps, scissors, poker, and the common balance.

In levers of the *second class* the applied force and the resistance are on the same side of the fulcrum, the former being farther from it than the latter. Thus the resistance is always greater than the applied force. This class includes the oar of a boat, a pair of nut-crackers, and a claw-hammer for extracting nails.

In levers of the *third class* the applied force and the resistance are on the same side of the fulcrum, the former being nearer to the fulcrum than the latter. The purpose of such a lever is a gain of displacement or of speed. This class includes the forearm, which is hinged at the elbow and acted on by the biceps at a distance of two or three inches from the elbow, a pair of tongs and the lever of a safety-valve for steam pressure.

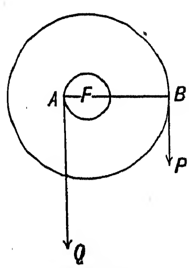


FIG. 71.—Wheel and axle.

143. The Wheel and Axle.—A straight lever cannot raise a weight higher above the fulcrum than the distance of the weight from the fulcrum. The apparatus called a "wheel and axle" acts on the same principle as a lever, but its range is not so limited (Fig. 71). It consists of a wheel of large radius, rigidly connected to an axle of smaller radius. The applied force, P, acts on a cord wrapped around the wheel, while the weight or resistance acts on a cord wrapped around the axle. The principle involved is that of a lever of the first class, the radius, R, of the wheel being the lever arm for the applied force, while the radius, r, of the axle is the lever arm of the resistance. Hence

$$\frac{Q}{P} = \frac{R}{r}$$

This formula may also be proved directly by equating the work done by P in one complete revolution, $2\pi RP$, to the work done against Q, $2\pi rQ$; also by taking moments about F.

The principle of the wheel and axle is applied in the pilot wheel and in the capstan, where the wheel is replaced by spokes in the axle, and in the winch, where there is but a single spoke, the crank arm.

In the above we have neglected friction, which is always considerable.

144. Differential Wheel and Axle.—To obtain a very high mechanical advantage, the wheel would have to be made very large, which would be inconvenient, or the axle would have to be made very small, which would greatly weaken it. To avoid these disadvantages, the axle is made in two parts, of different size, and the cord is wrapped in the same direction around both, as indicated in figure 72, the weight being carried by a pulley through which the cord passes.

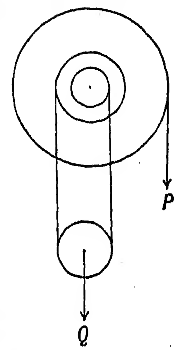


FIG. 72.—Differential wheel and axle.

Let the radius of the wheel be R, that of the large part of the axle r, and of the small part r'. The upward force on the pulley is twice the tension of the cord, and the downward force is Q, the weight of the pulley being neglected. Hence, by the principle of forces in equilibrium, the tension in the cord is $\frac{1}{2}Q$. In one revolution P does work $P \cdot 2\pi R$, and the tension of the cord, acting on the smaller part of the axle, does work, $\frac{1}{2}Q \cdot 2\pi r'$, while work $\frac{1}{2}Q \cdot 2\pi r$ is done against the tension in the cord, acting on the larger part of the axle. Hence

$$P \cdot 2\pi R + \tfrac{1}{2}Q \cdot 2\pi r' = \tfrac{1}{2}Q \cdot 2\pi r$$

$$\therefore \frac{Q}{P} = \frac{2R}{r - r'}$$

145. Pulleys.—The simplest pulley is a wheel for the purpose of changing the direction in which a force is applied. It consists of a wheel in a frame-work or block which is either fixed or free. If it is fixed, the direction of the force is changed, without any change in the magnitude of the force (Fig. 73a).

If it is free and the two parts of the cord are parallel (Fig. 73b), the tension in any part of the cord is (neglecting friction and the

weight of the cord) equal to the force applied at its free end. Hence, for equilibrium,

$$Q = 2P$$

If the weight of the pulley is not negligible, it may be included in Q. This formula is also readily found by the principle of energy; for each unit of length that Q moves P must move two.

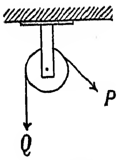


FIG. 73a.

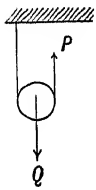


FIG. 73b.

146. Block and Tackle.—Several pulleys are frequently used in combination, so as to secure higher mechanical advantage. The most common arrangement is called the *block and tackle.* The pulleys are in two blocks, with several pulleys in each block. The fixed end of the cord may be attached to either the upper or the lower block; if to the former, there will be an equal number of pulleys in the two blocks, as in the figure; if to the latter, there will be one more pulley in the upper block. When the distance between the blocks is decreased by one unit of length, each branch of the cord in contact with the lower pulley must shorten one unit of length. Hence

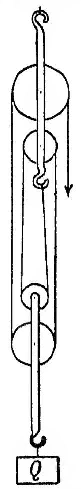


FIG. 74. Block and tackle.

$$Q = nP$$

where n is the number of branches of the cord at the lower block.

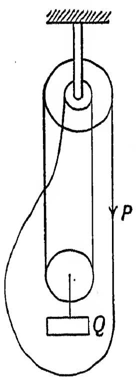


FIG. 75.

147. The Differential Pulley or Chain Hoist.—In this the upper block holds two pulleys, of different diameters, fixed rigidly to the same axis, while the lower block is replaced by a single pulley. An endless chain passes over the three pulleys, as shown in Fig. 75, and is prevented from slipping by teeth on the pulleys. This is essentially a modification of the differential wheel and axle, the wheel and the larger part of the axle having the same radius. The relation between P and Q, which may be worked out independently or may be obtained by putting $R = r$ in the formula of §144, is

$$\frac{Q}{P} = \frac{2r}{r - r'}$$

148. The Inclined Plane.—A force less than the weight of a body may suffice to draw the body up an inclined plane. Let P be the force and W the weight (Fig. 76a). Also let h be the height and l the length of the plane. When the body has been drawn up the whole length of the plane, the work done by P (neglecting friction) will be Pl and the work done against W, or the increase of potential energy, will be Wh. These must be equal. Hence

$$\frac{W}{P} = \frac{l}{h}$$

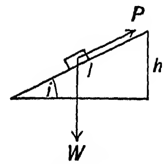


FIG. 76a.

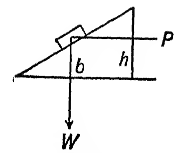


FIG. 76b.

This is essentially the same expression as already found (§52) by considering the component of W down the plane. If friction cannot be neglected the work done against it will be Fl, where F is the force of friction, and in the above equation P must be replaced by $(P - F)$.

If P act horizontally (Fig. 76b) the work done by P will be Pb. Hence (neglecting friction)

$$\frac{W}{P} = \frac{b}{h}$$

149. The Screw.—The *thread* of an ordinary screw makes a constant angle with the axis of the screw. If the thread of a vertical screw were supposed unwrapped, with its inclination kept constant, it would be an inclined line. The *pitch* of a screw is the distance, parallel to the axis of the screw, between consecutive turns of the thread. The pitch divided by the outer circumference is the tangent of the inclination of the thread to the axis of the screw.

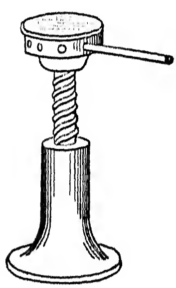


FIG. 77.—Jackscrew.

If a nut, carrying a heavy weight, be turned around a vertical screw, so that it ascends, the process will be similar to forcing a heavy body up an inclined plane by a horizontal force. In the jackscrew for raising heavy bodies the nut is fixed, while the screw is turned by a lever. The useful work performed by the screw in one turn is the product of the resistance it overcomes, Q, and the rise in one turn, which is the pitch h. The work done by the applied force in the same time is the product of the applied force, P, and the circumference $2\pi R$, of the end of the lever arm. Equating these

would give us a relation between P and Q if friction were negligible; but friction is, in general, so large as to make the relation of little value.

GRAVITATION

150. Law of Universal Gravitation.—Until the time of Newton, the weight of a body, or the measure of its tendency to fall to the earth, was generally regarded as an inherent property of matter that needed no further explanation. To Newton (and to some of his contemporaries) it occurred that the weight of a body on the surface of the earth is due to a force of attraction between the body and the earth, and that this attraction is only a particular case of a universal attraction between all bodies, no matter where situated Newton then sought to discover the law that such a force would have to follow to account for the facts, how it would have to depend on the masses of the bodies and their distance apart. Now it was not possible for him to change the distance between a body and the center of the earth by any except an exceedingly small fraction, and the force between two bodies of ordinary size on the surface of the earth was so small that it escaped detection until a much later date. Hence he turned his attention to the motion of the moon and the planets.

Before the time of Newton, Kepler had, by a very extensive and painstaking study of the motions of the planets, arrived at certain laws known as *Kepler's Laws*. These may be stated as follows:

1. Each planet moves in an ellipse, at one focus of which the sun is situated.

2. The areas swept over by a line joining a planet to the sun are proportional to the times.

3. The squares of the periods of revolution of the planets are proportional to the cubes of the major axes of the ellipses.

From these laws Newton showed that the motions of the planets could be accounted for on the supposition that, between each planet and the sun, there is a **force of attraction, proportional to the product of the masses and inversely as the square of their distances apart.**

Newton also showed that, if we suppose that there is a force according to this law between every two particles, a sphere that is either homogeneous or may be regarded as made up of shells,

each of which is homogeneous, will attract an outside body as if the sphere were concentrated at its center. The earth is very nearly such a sphere and must, therefore, according to the law of gravitation, attract (approximately) as if concentrated at its center.

151. Motion of the Moon.—As evidence for the law of gravitation, Newton showed that it correctly accounts for the motion of the moon. At the surface of the earth, a body is attracted by the earth, as if the latter were concentrated at its center. Now the radius of the earth is approximately 4,000 miles, and the average value of the acceleration of a falling body may be taken as 32.2 feet per sec.^{2} The distance of the moon from the earth, which is somewhat variable, may be taken as approximately 240,000 miles, or 60 times the radius of the earth. Hence, according to the law of gravitation, the acceleration of a body at the distance of the moon due to the earth's attraction should be $32.2/60^2$ or .00894 ft. per sec.^{2}

The acceleration, a, of the moon towards the earth (§33) equals v^2/R. The period of rotation of the moon, also slightly variable, is about 27 days, 8 hours. Calling this T, we have $v = (2\pi R/T)$. Hence $a = (4\pi^2 R)/T^2$, or, reducing R to feet and T to seconds, $a = .00896$. This value of a, calculated from the observed period of the moon, agrees as closely with the preceding value, deduced from the law of gravitation, as could be expected, when it is remembered that only approximate values for the various constants have been used. More elaborate calculations yield a satisfactory agreement between the two values of a.

152. Force of Gravitation Proportional to Mass.—According to the law of gravitation, the attraction between two bodies is proportional to their masses and is independent of the materials of which they consist. One proof of this was given by Galileo, when he dropped two cannon balls of different sizes from the leaning tower of Pisa and found that they reached the ground in very nearly the same time. Their accelerations being equal, the ratio of the force to the mass must, according to the second law of motion, be the same for both. Yet, in Galileo's experiment, the larger weight was slightly ahead of the smaller, and Galileo correctly explained this difference by remarking that the air-friction would be proportionately less on the larger body. In fact, because of this air-friction and the rapidity of the motion, it would be

difficult to give a very convincing proof of the law by means of bodies falling with the full acceleration due to gravity.

To avoid this difficulty, Newton experimented with a pendulum, the motion of which depends on gravity, but on a fraction only of the full force of gravity, namely, the component along the arc of vibration. The bob of the pendulum was a thin shell and into this he put in successive experiments different substances. In each case the same *weight*, as tested by weighing with a balance, was put into the box, and, since the force of air-friction on the box for the same amplitude of vibration would be the same, no matter what the contents of the box, it followed that, at a given inclination to the vertical, the force causing the motion would be always the same. He found that the time of vibration was always the same, no matter what the contents of the box, and hence the masses must also have been the same; that is, equal masses of different substances have equal weights. These experiments were afterward repeated by Bessel with much greater care and with the same result.

The above experiments prove that gravitation is not, like magnetic attraction, a force that depends on some quality of a body, other than its mass, that is, not a selective force but a general force. That it does not depend on any other physical condition, such as temperature, or on any chemical condition, such as molecular combination, has also been shown by most careful weighing. A third body, placed between two bodies, has not the least effect in shielding them from their mutual attraction. The fact that a lump of gold, when hammered out into an exceedingly thin sheet, suffers no change of weight shows that the weight of a body does not depend on its form, that gravity acts on the particles whether surrounded by other particles of the same kind or not.

153. The Constant of Gravitation.—The law of gravitation may be stated as a formula,

$$F = G\frac{mm'}{r^2}$$

where G is a constant, called *the constant of gravitation.* To find the magnitude of G, it is necessary to measure F in some case where m, m' and r are all known. This was first done by Henry Cavendish in 1797–8, and the experiment, usually called the *Cavendish experiment*, has been repeated many times since, with increasing

care and accuracy. Cavendish suspended two balls, A and B (Fig. 78), from the ends of a long, light, horizontal rod, which was supported by a long, fine, vertical wire, attached to the middle, C, of the rod. On opposite sides, horizontally, of the balls, and at known equal distances, he placed two large spheres of lead, P and Q. The attraction between each ball and the adjacent large sphere had a moment about C that produced a twist of the supporting wire. When the spheres were in the position P_1Q_1, the twist was in one direction, and when they were in the position P_2Q_2 the twist was in the opposite direction. To deduce the force of attraction from the magnitude of the twist, the constant of torsion of the wire (§124) had to be found by timing vibrations of the wire when the spheres were removed to positions where they had no influence on the vibrations of AB. Thus F, m, m', and r were found, and, when they were substituted in the above formula, the value of G was obtained.

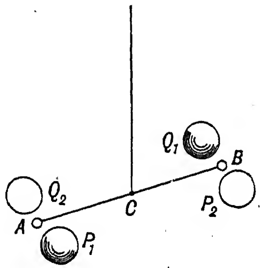


FIG. 78.—Principle of the Cavendish experiment.

In more recent work the apparatus has been greatly improved. The greatest improvement has been in the substitution of very fine quartz thread for the wire. This also permitted the apparatus to be greatly reduced in size, so that, whereas AB in Cavendish's experiment was 6 ft. long, in Boys' apparatus it was only 0.9 inch, and the masses A, B, and P, Q, were also greatly reduced in size. The value obtained for G (using c.g.s. units) was 6.6579×10^{-8}; this is, therefore, the force in dynes of the attraction between spheres of one gram each at a distance of 1 cm. between their centers. When it is remembered that a dyne is about the weight of a milligram, it is seen that the force measured in the above experiments was exceedingly small; hence the difficulty of the experiment.

154. The Mean Density of the Earth.—The determination of G made it possible to calculate the mass of the earth (hence Cavendish is sometimes said to have been the first to "weigh the earth"). For if m', in the formula for the law of gravitation, be put equal to one gram, and m and r be taken as, respectively, the mass and radius of the earth, F will be the force of attraction between the earth and a body of 1 gm. mass, and this is, as we know, 980 dynes. Thus the formula gives us the value of m, the mass of the earth,

which is found to be 5.97×10^{27} gms. This figure is so large as to convey no distinct meaning, but a different way of stating the result will be more easily comprehended. The density of a homogeneous body, such as water, is its mass per unit volume, and, when the density of a body is not everywhere the same, we may speak of its *mean density*, or its whole mass divided by its whole volume. Thus, to get the mean density of the earth, we divide its whole mass, as given above, by its whole volume. The result is 5.527, that is to say, on the average the earth is 5.527 times as dense as water. More recent work seems to show that this figure should be about 5.515. It is remarkable that Newton, reasoning from the very slight evidence available in his time, supposed the mean density of the earth to be between 5 and 6.

155. The Tides.—At any place on the shore of an ocean, the level of the water rises to a maximum and falls to a minimum once in about every twelve hours and 25 minutes. These risings and fallings are called the *tides*. They are due to the forces of attraction that the moon and the sun exert on the water on the surface of the earth and to the rotation of the earth. The complete explanation of their action is extremely difficult, owing to the irregularities of the continents and to other causes.

It has been found that the solid earth itself yields, to the extent of about 9 inches at a maximum, to the tide-raising force of the sun and moon.

PROPERTIES OF MATTER

GENERAL PROPERTIES

156. Internal Forces in Matter.—In the preceding chapters on the principles of Mechanics we have been studying forces between bodies. The forces that the particles of a body exert on one another did not need to be considered, for they cancelled out when the action of the body as a whole was considered. We shall now consider some of the properties of matter that depend on the forces between particles. It will be seen that some of the relations between these properties are not as definite and calculable as the relations between the quantities studied in Mechanics. This is chiefly because the ultimate particles of a body are so small that they cannot be studied separately. In fact, we can only infer their existence and relations

from the properties they exhibit in the large aggregations of particles that we call bodies.

157. The Three States of Matter.—Following popular language, we classify bodies as solids and fluids. The characteristic of a *solid* is that it has a definite shape, which it does not readily relinquish, while a *fluid* flows easily, or suffers a change of shape in response to the smallest distorting influence. (Some bodies, such as pitch, are in a borderland between the two classes.) The particles of a solid are held in definite positions by the forces between them, but each particle has some freedom to vibrate about its mean position (see §161).

Fluids are divided into *liquids* and *gases*. The peculiarity of a liquid is that, while it readily flows, it has a definite volume, which it does not readily change. A gas yields to the smallest force exerted to change its volume, in other words, it has no definite volume of its own but takes the volume of the containing vessel, however large. The particles of a liquid are close together and attract each other with powerful forces. These forces react strongly against outside forces that tend to change the mean distance between the particles, but they are such as to permit sliding motions of the particles. The particles of a gas are practically separate bodies, flying in space and exerting no appreciable forces on one another, except at impact of particle on particle.

158. Elements and Compounds.—In innumerable cases two or more substances coalesce to form a new substance, which may be so distinct in all its properties that nothing, apparently, remains to suggest the constituents from which it was formed. Thus, two substances, oxygen and hydrogen, gaseous under ordinary conditions, combine to form a liquid, water. Harmless substances may, on combination, form deadly poisons or explosives. Substances that may be made from constituents that have properties distinct from the resultant are called *compounds*.

Conversely, compounds may be divided up into constituents differing widely from the original substance, and these constituents may be themselves capable of being resolved into other constituents. But there are 92 substances that are classified as chemical *elements*, though it is now known that a few of them decompose spontaneously or can be disintegrated by special methods which we shall consider later.

159. Molecules and Atoms.—Many facts, chiefly such as are more closely studied in chemistry, justify the belief that (1) an element consists of very small particles called *atoms*, (2) the atoms in one elementary body are different from those of any other elementary body, (3) these atoms are usually combined in similar groups called *molecules* (in some substances the atom and the molecule are identical). There is also reason to believe that in many substances, espe-

cially liquids and solids, molecules are frequently combined to form groups or molecular aggregates of two or more molecules each.

Atoms are extremely small and will probably never be separately visible, however much optical instruments may be improved but some molecules, e.g. of rubber, can be made visible. In a cubic centimeter of a gas under ordinary conditions there are about 2.7×10^{19} molecules.

There is good reason to believe that atoms contain still smaller parts called *electrons*, which may pass from atom to atom, and are sometimes entirely separated from atoms. The properties of these explain many of the phenomena of light and electricity.

The atoms of an element are not all of the same weight, though they are identical in chemical properties. The different kinds of atoms of an element are called *isotopes* of the element. For example, zinc has five isotopes, the atomic weights of which are 64, 66, 67, 68, 70. In such cases the atomic weight of the element is the mean of the atomic weights of the isotopes, allowance being made for the proportions in which these are mixed.

160. Intermolecular Forces.—It is evident from the great forces necessary to pull a solid body apart that there are comparatively great forces between molecules; but the ease with which a brittle body falls apart on the appearance of a slight crack shows that the forces are not appreciable unless the molecules are very close together. The latter point is also shown by the fact that a body reduced to powder, *e.g.*, the graphite of which lead pencils are made, can only be changed back into a compact solid by very great pressure.

Roughly speaking, it may be said that the force of molecular attraction in water is inappreciable at distances greater than about 10^{-7} cm. The magnitude and the range of the intermolecular forces are, of course, different for different substances, and the characteristic properties of different substances depend to some extent on these differences.

161. Kinetic Theory of Matter.—There is very strong evidence that the particles of which bodies are made up are not at rest. Thus two different gases, contained in two different vessels, mix with great rapidity when the vessels are put in communication. This process is called *diffusion*. Liquids will also diffuse into one another (even, to a slight extent, liquids like oil and water), though much more slowly than gases, because of the greater closeness of the particles and the frequent changes of direction of motion of a particle, produced by impact on other particles. Some solids show, by diffusion, that their particles are not at rest; thus, when a small block of gold is placed on a block of lead, with planed surfaces in

close contact, after the lapse of some weeks it is possible to detect particles of gold which have wandered into the lead and *vice versa.* There are many other reasons for believing that the particles of matter are in all cases in motion. This condition is expressed by the phrase: the *kinetic constitution of matter.*

162. Density and Specific Gravity.—The *density* of a body is its *mass per unit volume.* If the masses of all equal volumes of the body are the same, the density is *uniform* and equal to the mass in any unit of volume. If the masses of equal volumes are not the same, the density is not uniform. The *mean density* in any particular volume of the body is the mass in that volume divided by the volume. The *density at a point* is the mean density in a small volume enclosing the point when the volume is supposed to be decreased without limit.

The measure of the density of a body depends, of course, on the units of mass and volume employed. When c.g.s. units are employed, density is the number of gms. per c.c. In these units the density of water at 4°C. is very nearly unity, since the gram was originally intended to be the mass of 1 c.c. of water at 4°C. In British units the density of a body is the number of lbs. per cu. ft. of a body. In these units the density of water is 62.4, since that is the number of lbs. in a cu. ft. of water.

The *specific gravity* of a body is the ratio of its density to that of some standard substance. The standard usually employed is water at 4°C. Thus, if ρ be the density of the body and ρ' that of water at 4°C., the specific gravity of the body is ρ/ρ'. Now in c.g.s. units ρ' is very nearly unity. Hence in these units density and specific gravity are numerically equal. But in British units, since ρ' is 62.4, the specific gravity of a body is its density divided by 62.4. Hence *in British units*

$$\textit{density (lbs. per. cu. ft.)} = \textit{specific gravity} \times 62.4$$

Table of Densities (gms. per c.c.)

Aluminum	2.60	Iron (about)	7.60
Brass (about)	8.50	Lead	11.37
Copper	8.92	Platinum	21.50
Gold	19.32	Silver	10.53
Ice	.917	Air at 0° and 1 atmo	.001293
Alcohol at 20°	.789	H at 0° and 1 atmo	.00008988
Ether at 20°	.715	N at 0° and 1 atmo	.001256
Mercury at 20°	13.55	O at 0° and 1 atmo	.001430
Sea Water (about)	1.026	Heat 0° and 1 atmo	.0001785

PROPERTIES OF SOLIDS

163. Homogeneity and Isotropy.—A homogeneous body is one which has the same properties *at all points*, so that small spheres of equal radii, cut out of different parts of the body, would be identical in properties. Many crystals are nearly perfectly homogeneous, and so, too, is good glass, such as plate glass or the glass of lenses. Many other bodies are approximately homogeneous, for example, most metals, wood, stones, etc.

An **isotropic** body is one which has, at any point, the same properties *in all directions*, so that if, at any point, a sphere were cut out there would be nothing in the properties of the sphere to indicate the original direction of any diameter. All liquids and gases are isotropic under ordinary conditions, but many substances, such as crystals, woods, and drawn metals, are distinctly non-isotropic.

164. Elasticity.—When the shape or volume of a solid is changed by the application of some force, there is, in most cases, a tendency to return to the original shape or volume, when this force is removed. This *tendency to recover from distortion* is called *elasticity.* It is one of the most important properties of a solid, since the usefulness of many bodies, such as springs, musical instruments, etc., depends on the extent to which they possess it. It is therefore a property that has been very extensively studied.

165. Strain.—**Any change of shape or of volume or of both is called a strain.** Thus the bending of a beam, the twisting of a rod, the compression of a liquid or a gas into a smaller volume are strains. The term strain is a geometrical one, and its definition contains no reference to force or energy, although, as we shall see, force and energy are present when a body is in a state of strain.

A strain that consists in a *change of shape only,* without any change of volume, is called a *shear.* The strain of a moderately twisted wire or rod is a shear.

A strain that consists in a *change of volume* only, without any change of shape, has not received any special name, but we may for brevity call it a *volume-strain.* Such, for example, is the strain of a sphere of cork or of any isotropic body, when it is placed in a fluid in a closed vessel and is subjected to great pressure.

While, for simplicity, we have first enumerated strains in which either volume or shape alone changes, strains that involve changes

of both are more common. Thus the stretching of a wire, the compression of a pillar, the bending of a beam, etc., are strains of both volume and shape. A body is said to be homogeneously strained, or the strain is described as *homogeneous*, when the nature and magnitude of the strain is the same at all points in the body. Thus, when a wire is stretched or a rod compressed, and when a liquid or gas is subjected to pressure, the strain is homogeneous. But when a wire or rod is twisted, the strain is greatest at the surface and least at the center, and when a beam is bent, there is a stretching on the convex side and a compression on the concave side, and the strain is heterogeneous.

166. Stress.—When a body is in a state of strain, owing to the action of external forces on it, there are **internal forces between contiguous parts of the body,** in addition to whatever internal forces there may have been before the strain occurred. If, at a point, a dividing plane be imagined, the part of the body on one side will act with a certain force on the part on the other side, and the latter will react with an equal and opposite force. These two forces, the action and the reaction, constitute a *stress*. In some cases the force is perpendicular to the imaginary dividing plane and may be called a *thrust*. In other cases it is parallel to the dividing plane. In any case *the magnitude of the stress is the force per unit area* of such an imaginary dividing plane.

The terms *homogeneous* and *heterogeneous* apply to stress just as to strain. In some cases, for example in the stretch of a wire by an attached weight, the stress in a body is equal to the external force per unit area that acts on the body and produces the strain, and in such cases this external force per unit of area is often called the stress. In other cases, as, for example, in the bending of a beam by a weight that acts at some point, the stress does not bear a simple relation to the external force, and we must take care to distinguish them.

167. The Measure of a Strain and of a Stress.—A strain that consists in a change of volume only is measured by *the proportion in which the volume is changed.* If the strain is homogeneous, the measure may be taken as the change in unit volume, or, if a volume V becomes $(V \pm v)$, the measure of the strain is v/V. If the strain is not homogeneous, the measure of the strain at any particular point is the value of v/V at the point, when V is taken as the volume of an unlimitedly small portion surrounding the point.

To produce this change of volume, force must be applied to the surface of the solid, in the form of either a push or a pull; and, inside the body, each part will push or pull on each neighboring part. The amount of this push or pull per unit area is the measure of the stress. Instead of the words push and pull we shall hereafter use the more general term thrust, a push being a positive thrust, and a pull a negative thrust.

The measure of a *shear* will be most readily understood by considering the simplest way in which a shear may be produced. Consider, for example, a rectangular block of rubber (or stiff jelly) between two boards to which it adheres. Let $PQRS$ be one rectangular face and PQ, RS the edges of the boards. Apply to the boards equal and opposite forces, parallel to them and to the face $PQRS$. The face $PQRS$ is changed to the form P_1Q_1RS. Each section of the block parallel to the boards moves parallel to itself a distance proportional to its distance from RS. Each of the right angles of $PQRS$ is changed by the same amount, say θ, and this change is taken as the measure of the shear. When θ is small, as it is in most practical cases, the magnitude of the angle θ, in radian measurement, is P_1P/PS or *the relative displacement of two planes unit distance apart.*

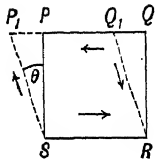


FIG. 79.—Shear and shearing stress.

If an imaginary plane be supposed drawn anywhere in the block, parallel to the boards, the part on one side of this plane will exert a tangential force on the part on the other side, and this force will equal the force applied to the boards. The *magnitude of the force per unit area* of this plane is the measure of the shearing stress.

While we have referred only to the forces parallel to PQ and RS, it is clear that the shear cannot be produced without other forces being applied to the block. If only the two forces described were applied, the block would not be at rest but in rotation, since the two constitute a couple. The effect is readily perceived when the attempt is made to apply the two opposite forces. It is, in fact, necessary also to apply other forces, forming an opposite counterbalancing couple, say along SP_1 and Q_1R. The effect of all four forces is to produce a stretch along RP_1 and a compression along Q_1S, and the proportional stretch is equal to the proportional compression, since there is no change of volume.

168. Hooke's Law.—When any body is strained beyond a certain amount and then released, it fails to return completely to its original form and volume, or it retains a *permanent set*. The largest strain of any kind which a body may undergo and still completely recover from when released is called *the limit of elasticity* for that form of strain, and the corresponding stress is called the *limiting stress*. The limit of elasticity is, of course, widely different for different substances. Thus, rubber may be greatly extended and yet recover, while the limit for glass and ivory is very small. (Cases in which the limit is somewhat indefinite will be considered later.)

Within the limit of elasticity, a simple law, first stated by Hooke in 1676 and known as *Hooke's law*, holds, namely, **stress is proportional to strain.** (Hooke's statement in Latin was "Ut tensio sic vis.") Hooke illustrated his law by various cases of strain, such as the stretching of a spiral spring and of a wire, the bending of a beam, and the twisting of a wire.

169. Moduli of Elasticity.—While elasticity has already been defined as the tendency of a body to recover its shape or volume when distorted, the definition is purely qualitative and affords no means of assigning a numerical value to the property. A quantitative definition of the elasticity of a substance for any form of strain follows from Hooke's law. **The measure or modulus of elasticity is the ratio of the magnitude of the stress to that of the accompanying strain,** this ratio being a constant within the limits of elasticity. As there is a great variety of forms of strain, there is a correspondingly large number of moduli of elasticity for any substance; but only three of these are important enough to be enumerated here.

When the strain is one of volume only, the elasticity is called *elasticity of volume*. The modulus of elasticity of volume or the *bulk modulus*, as it is frequently called, is the ratio of the stress, or the thrust per unit area, p, to the change of volume per unit volume produced. The bulk modulus of a substance is usually denoted by k. Hence, if a volume V is decreased by an amount v,

$$k = p \div \frac{v}{V} = \frac{pV}{v}$$

The reciprocal of the bulk modulus is called the coefficient of *compressibility* of the substance. It means the ratio of the proportional compression to the thrust per unit area.

When the strain is a shear, the modulus of elasticity, called the *shear modulus*, or often the *simple rigidity*, is the ratio of the shearing stress to the shear. Denoting the shearing stress by T, the shear corresponding to T by θ, and the shear modulus by n,

$$n = \frac{T}{\theta}$$

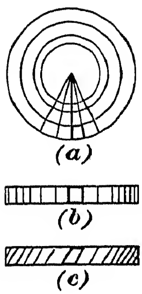


FIG. 80.—Shear of a small cube in a twisted wire.

170. Torsion.—When a wire or rod of homogeneous isotropic material is twisted (Fig. 80), we may imagine the whole length divided into transverse slices, of equal thickness, by planes perpendicular to the axis. Each such slice will be rotated about the axis, to an extent proportional to its distance from the fixed end. Moreover, one face of each slice (the one farthest from the fixed end) will be rotated more than the other. Let us now suppose that each slice is very thin, and that it is divided up before twisting into very small cubes (or nearly cubes) by a series of imaginary planes through the axis, intersected by concentric cylinders. Thus each cube will have four edges parallel to the axis, four others in the direction of radii, while the remaining four will be short and practically straight arcs of circles. After the twist each cube will have a strain like the cube of rubber in §167. Hence the strain is a shear, but, since the strain of each cube will be proportional to its distance from the axis, the strain is not homogeneous.

171. Young's Modulus.—A very frequent form of strain is that of a uniform wire or rod that is clamped at one end and is acted on by a longitudinal force at the other end. Such a strain is called a *stretch*. Any short part of the wire is extended in the same proportion as the whole wire. The measure of the stretch is the extension per unit length, or, denoting the unstretched length of the wire by L and the total extension by l, the stretch is l/L. The measure of the stress is the pull per unit of cross-sectional area. If F is the whole force applied to one end and A the cross-sectional area of the wire and the weight of the wire is neglected, the pull per unit area anywhere in the rod, due to the force F, is F/A, which is, therefore, the measure of the stress. Young's modulus, which we may denote by M, is, therefore, $(F/A) \div (l/L)$ or

$$M = \frac{FL}{Al}$$

For some common materials the average values of k, n, and M in dynes per cm.[2] are as follows:

	k.	n.	M.
Copper	17×10^{11}	4×10^{11}	11×10^{11}
Glass	4×10^{11}	2×10^{11}	6×10^{11}
Iron (wrought)	15×10^{11}	7×10^{11}	19×10^{11}
Lead	4×10^{11}	2×10^{11}	1×10^{11}
Steel	17×10^{11}	8×10^{11}	23×10^{11}

172. Volume Changes when a Wire is Stretched.—When a wire or rod is stretched there is obviously a change of shape in every part of the wire or rod, for the length is increased, while the cross-section is decreased. Whether a change of volume also occurs can only be determined by experiment. If the cross-section diminishes in the same proportion as that in which the length increases, there is no change of volume; whereas, if the proportion in which the length increases exceeds that in which the cross-section diminishes, there is an increase of volume. Careful experiment shows that in all cases there is an increase of volume; but in some substances, *e.g.*, india rubber, the proportional change of volume is very small.

173. Flexure.—A very common strain, closely related to stretching, is that of a plank, supported at both ends and carrying a load at the middle, or supported at the middle and loaded at each end, or clamped horizontally at one end and loaded at the other end (Fig. 81). A little consideration will make it clear that, in these cases, we have to do with stretches and shortenings, like those already discussed. If we suppose the plank divided into a large number of longitudinal strips, the strips on the convex side are stretched by the bending, while those on the concave side are shortened. There must, of course, be an intermediate surface, where there is neither stretch nor compression, and this surface is called the *neutral surface*. The extension or compression of any strip is proportional to its distance from the neutral surface. Thus the strain, while not homogeneous, is everywhere of the nature of an extension or a compression, and Young's modulus is the only modulus involved. If a bar of length l, breadth b, and depth d cms. be supported at both ends and be subjected to a perpendicular force of F dynes at the middle, the depression produced is $Fl^3/4Mbd^3$.

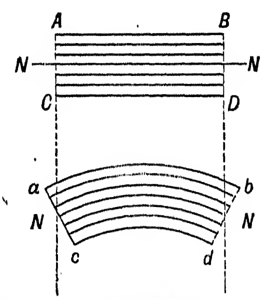


FIG. 81.—Bending of a beam (exaggerated).

174. Direct Impact of Elastic Bodies.—When two bodies in motion collide, each exerts a momentary force on the other, and

each, therefore, suffers a change of velocity. The result is difficult to calculate, except in certain simple cases.

When the bodies are uniform spheres and are moving before impact along the line joining their centers, the result can be calculated. Let the masses be m and m' and let the velocities before impact be u and u', both being in the positive direction, with $u > u'$. After the impact, m' will be moving faster than m. Let the velocity of m after impact be v, and let that of m' be v'. Then $v' > v$ (unless the substances be so soft that v and v' are equal). During the short time of contact, each body exerts a force on the other, and, by the Third Law of Motion, these forces are, at any moment, equal and opposite. These forces also act for the same length of time and must, therefore, produce equal and opposite changes of momentum. Hence *the total momentum after impact equals the total momentum before impact* (§47), or

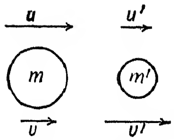


FIG. 82.—Motion of spheres before and after impact.

$$mv + m'v' = mu + m'u' \tag{1}$$

If the problem be to find the velocities after impact, this equation will not suffice, since it contains two unknown quantities, v and v'. A second relation between v and v' was discovered experimentally by Newton. He found that, for given materials, the *ratio of the speed of separation,* $(v' - v)$, *to the speed of approach,* $(u - u')$, *is a constant,* which is (at least very nearly) independent of the masses and velocities of the bodies and depends only on their materials and on the direction of the grain, if they are not isotropic. This constant ratio is called the *coefficient of impact* (often called coefficient of restitution). Denoting it by e, we have

$$\frac{v' - v}{u - u'} = e$$

$$v - v' = -e(u - u') \tag{2}$$

From (1) and (2) v and v' can be calculated. For simplicity in establishing these equations, we choose the case in which all the velocities are in the positive direction, but they are algebraic equations, applicable to all cases. When numerical values of velocities are substituted, *all in one direction must be taken as positive and those in the opposite direction as negative.*

Some simple deductions can readily be drawn. When e is zero, as it very nearly is for such soft substances as putty and lead, we see from (2) that v and v' are equal, or the bodies do not separate after impact.

If the masses of the spheres be equal and $e = 1$, the spheres will, on impact, exchange velocities. For in this case the two equations become

$$v + v' = u + u'$$
$$v - v' = -u + u'$$

Hence $v = u'$ and $v' = u$, which proves the statement.

If one of the bodies, say m', be of very great mass compared with the other and be initially at rest, its velocity after impact will be very small. Putting both u' and v' equal to zero in (2) we get

$$v = -eu$$

This is the case when a small ball is dropped on a very large block. Let the height of fall be H and the height of rebound h. Then, taking downward as positive, $u = \sqrt{2gH}$ and $v = -\sqrt{2gh}$. Hence by the last result

$$e = \sqrt{\frac{h}{H}}$$

This affords a simple experimental method for finding e.

175. Oblique Impact of Smooth Spheres.—The impact of two spheres is described as oblique when the spheres are not moving before impact in the direction of the line through their centers. The lines of motion of the centers before impact may be in one plane, as when two equal balls rolling on a plane surface impinge, or these lines may be in different planes. In either case we may resolve the velocity of each ball before impact into two components, one in the direction of the line through the centers at the moment of impact, the other in a direction perpendicular to that line. Only the first component will be affected by the impact, since (if the spheres be supposed frictionless) the only force will be in the line of the centers. The change in this component may be calculated as in the case of direct impact: then by compounding this component for each sphere after impact with the unchanged component we can find the motion of each sphere after impact.

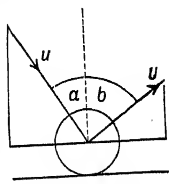


FIG. 83.—Impact on a fixed surface.

When a smooth ball impinges obliquely with a velocity u on a fixed surface in a direction making an angle a with the normal, its component velocity parallel to the surface is $u \sin a$ and that perpendicular to the surface is $u \cos a$. If it rebounds with a velocity v in a

direction making an angle b with the normal, the components become $v \sin b$ and $v \cos b$. The component parallel to the surface is not changed, while that perpendicular to the surface is changed in the ratio e:1. Hence, as regards the magnitudes of the velocities

$$v \sin b = u \sin a, \qquad v \cos b = eu \cos a.$$

Dividing corresponding sides we get

$$\tan b = u\frac{\tan a}{e}$$

Thus the direction of rebound is more nearly parallel to the surface than that of impact. This is the basis of a method that has been employed for finding e.

176. Loss of Energy on Impact.—The kinetic energy of twc smooth spheres before impact and that after impact can be calcu- lated from their masses and velocities. The total kinetic energy o two bodies is less after impact than before (except when e is unity) and other forms of energy, such as heat and sound, are produced.

177. Vibration of Elastic Bodies.—When a body is strainec within the limit of elasticity, the internal stresses tend to restor the body to its original condition. When quickly released from th external deforming force, the body vibrates, and, since the restorin forces are at each stage proportional to the distortion, the vibra tions are simple harmonic vibrations of a constant period. This for instance, is the case when a rod, firmly clamped at one end, i bent and released. When the vibrations are sufficiently rapic as is the case of the prongs of a steel tuning fork, sound is producec and the ear can test the constancy of the period of vibration by th steadiness of the pitch; the vibrations gradually die down, that i the extent of the maximum strain in each vibration decreases, ye the period remains unchanged, showing that, within the limits c vibration, the stress is, so far as the delicate sense of hearing ca detect, accurately proportional to the strain. A tuning fork ca be made of any metal, of wood or other solid substance; and, whi the sound may in many cases be weak and short-lived, the stead ness of pitch while it lasts is an excellent proof of Hooke's law.

178. Strain beyond the Elastic Limit.—As an illustration what happens when a substance is strained beyond the elastic limi that is, beyond the range in which Hooke's law holds, we shall co sider the stretching of a wire. When a force that stretches beyond the limit is applied to it and this force is steadily increase it elongates in greater proportion for each successive equal increa

of the force. As the force is increased, at a certain strain, called the *yield point*, a very rapid increase of strain sets in at some point of the wire, and the strain at that point continues to increase, even if the force is not increased, until at last the specimen "necks in" and breaks. Beyond the yield point the substance flows much like a very viscous liquid. If, during this process, the force be diminished somewhat, the strain will still continue to increase, but at a diminished rate; and, when the force is diminished sufficiently, the strain ceases to increase before breaking occurs. If, at this stage, the applied force be removed entirely, the wire will contract somewhat, but a large permanent set will remain. The wire will then act like a different wire with a new elastic limit.

179. Elastic After-effects.—From strain within the elastic limit, the strained material completely recovers in the course of time and there is no permanent set; but frequently the immediate recovery on removal of the force is not complete, and there remains a small temporary set, from which the material only slowly recovers. This slow recovery from temporary set is called an *elastic after-effect.* It is shown by rubber and glass and other substances which consist of mixtures of diverse molecules; but crystals and quartz threads do not show it.

It is readily demonstrated by clamping both ends of a thick rubber tube and attaching a small mirror to the middle to reflect a beam of light on a scale. Such an arrangement will show a double after-effect, due to successive twists in opposite directions.

180. Fatigue of Elasticity.—The vibrations of a torsional pendulum are maintained by the elasticity of the wire; they slowly die away, owing to air resistance and internal friction in the wire. If the pendulum be, by some means, kept vibrating a long time and then released, the vibrations will die away more rapidly than before, as if the elasticity has become somewhat exhausted by prolonged exercise. This fatigue will persist for a long time, but the wire will promptly recover after being heated to about 100°C.

181. Hardness and Other Properties.—Molecular forces give rise to a number of properties of solids that have not yet been defined in any precise way, such as malleability, ductility, plasticity, and friability. The property called *hardness* deserves special mention, as it is of importance in engineering and in mineralogy. For engineering purposes it is usually tested by pressing a very hard steel ball against the substance to be tested, the ratio of the force

to the area of the spherical indentation being taken as the index of hardness (Brinell hardness numeral). When tested in this way, hardness is closely allied to tensile strength. Another test of hardness, used for brick or stone, is abrasion. To the mineralogist hardness means scratching power. Ten substances, beginning with talc and ending with diamond, are arranged in a numbered series so that each will scratch the one preceding it in the list but not the one that follows. A substance that will scratch one in the list but not the next is regarded as having a hardness between these two, and a corresponding number is assigned to it. This is merely a method of classifying. Thus hardness is a property that varies according to the way in which it is measured—a principle that is sometimes applicable to other properties.

PROPERTIES OF FLUIDS

182. A fluid is distinguished from a solid by the absence of permanent resistance to forces tending to produce a change of shape;

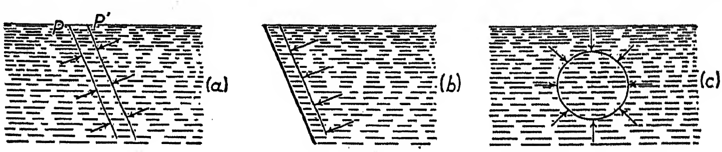


FIG. 84.

that is to say, **the shear modulus of a fluid is zero.** In this respect all fluids agree; they also agree in having weight and inertia. Because of agreement in these respects there are certain properties common to all fluids.

In certain other respects liquids and gases differ considerably. These differences are due to the fact that, while the particles of liquids are comparatively close together and attract one another with very considerable forces, the particles of gases are so far apart that the forces between them are usually negligible (except at impact). Properties in which liquids and gases differ will be treated in separate chapters.

183. Forces in a Fluid at Rest.—In a fluid at rest every part of the fluid is in equilibrium. Consider a thin layer of the fluid between two parallel planes, P and P' (Fig. 84a). The forces acting on it are its weight, w, and the forces, F and F', exerted on its faces by the adjacent fluid. If P and P' be taken sufficiently close together, the mass of the layer will be so small that its weight will

be negligible, compared with F and F'. Hence F and F' are forces in equilibrium, and they are therefore equal and opposite. Moreover, they must be perpendicular to the layer. If they were not, they would have components parallel to the layer, and these components would constitute a shearing stress, to which the fluid would yield, since its shear modulus is zero. To put the matter more briefly, since the shear modulus of a fluid is zero **the only form of stress possible in a fluid at rest is a normal stress.**

Similar reasoning can be applied to a thin layer of a fluid in contact with the surface of a solid (Fig. 84*b*). The force exerted by a fluid on the surface of a solid is at right angles to the surface, and the force exerted by the solid on the fluid is also at right angles to the surface of contact.

The force normal to the surface, that is exerted by a fluid is either a push or a pull. Usually it is a push, and it is then called a *thrust*, but in some exceptional circumstances it may be a pull, which we may regard as a negative thrust. While the term thrust is applied primarily to a force exerted on a bounding surface, it is convenient to extend it to any imaginary surface in a fluid; the force that each part of the fluid exerts on the other part, across the imaginary dividing surface, is called a thrust.

184. Pressure in a Fluid.—The *thrust per unit area* at a point in a fluid is called the *pressure* at the point. It is to be noted that, in the definition, the inclination of the area is not mentioned. We can justify this omission by considering the equilibrium of a small sphere of the fluid, having the point in question as center (Fig. 84*c*). If the surface of the sphere be supposed to be divided into a large number of small equal parts, the thrust on each of these will be toward the center. If these thrusts were not equal, the sphere would be deformed, that is, the fluid would not be at rest. If the radius of the sphere be now decreased toward zero, the thrusts per unit area will ultimately become the pressures in different directions, and these are therefore equal. We have, it is true, neglected the weight of the sphere, but, since it is proportional to the cube of the radius, while the thrust on each small part of the surface is proportional to the square of the radius, the weight becomes negligible when the radius approaches zero. The statement that pressure at a point in a fluid is the same in all directions, is equivalent to saying that pressure is a *scalar* quantity, that is, a quantity that is independent of direction (§15).

The pressure in a fluid may be traced to different causes. The *weight* of the upper layers of the fluid is sustained by the lower layers, and this produces pressure. When a fluid completely fills a vessel, *forces applied to the surface* cause pressure in the fluid. A liquid exposed to the atmosphere or any gas receives pressure from the gas. *Surface tension* is another cause of fluid pressure, and this will be considered later. Molecular attraction also produces pressure, but, for a reason stated later (§206), this is not taken into account at present.

185. Pressure at Different Points in a Fluid.—(1) Let P and Q be two points in a fluid at rest, the positions of the points being such that the straight line PQ is horizontal and wholly in the fluid. Consider the forces acting on a cylinder of the fluid, described about PQ as axis. The thrusts on the curved surface of the cylinder have no components in the direction of the axis. Hence, for equilibrium, the thrusts on the ends must be equal and opposite; and, since the ends are of the same area, the thrusts per unit area on the ends must be equal. If, now, the radius of the cylinder be supposed indefinitely decreased, the thrusts per unit area on the ends become the pressures at P and Q, which must, therefore, be equal. Hence the pressure in any direction at P equals the pressure in any direction at Q.

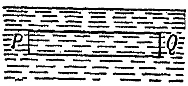


FIG. 85.

(2) Let P and Q be two points in a vertical line that is wholly in a fluid of density ρ. Consider the forces acting vertically on a cylinder of the fluid having PQ as axis and of unit, 1 sq. cm., cross-section. If the depth of Q below P be h cms. the volume of the cylinder will be h c.c., its mass will be $h\rho$ gms. and its weight $h\rho g$ dynes. If p_1 be the pressure in dynes per sq. cm. at P and p_2 that at Q, the thrust downward at P will be p_1; and that upward at Q will be p_2. Hence for equilibrium

$$p_2 - p_1 = h\rho g$$

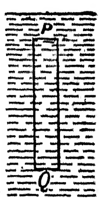


FIG. 86.

(3) Let P and Q be any two points in the fluid. No matter what the shape of the containing vessel, P and Q can be connected by a broken line, made up of vertical and horizontal steps (Fig. 87). Along the zigzag path from P to Q, there is a difference of pressure $h'\rho g$ for each vertical step of length h', while, for each horizontal step, there is no change of pressure. Hence

the difference of pressure between P and Q is $g\rho \times$ (the algebraic sum of the vertical steps) or, if the difference of level of P and Q be h, the difference of pressure is $h\rho g$.

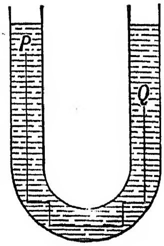


FIG. 87.

186. Pressure in a Gas.—Since the density of a gas is comparatively small, the difference of the pressures at two points is usually so slight as to be negligible; but this is not the case if h be very great. Thus in a vessel containing gas the pressure may be regarded as everywhere the same; but the pressure of the air varies greatly as we ascend to great heights in the atmosphere or descend to great depths in a mine.

187. Units Employed in Calculating Fluid Pressure.—In establishing the formula for difference of pressure at different depths in a fluid, namely,

$$p_2 - p_1 = h\rho g$$

it has been supposed that h is in cms, ρ in gms. per c.c., g in cm. per sec.^{2} (about 980), and p_1 and p_2 in dynes per sq. cm. A dyne per sq. cm. is sometimes called a *bar*.

When the values of p_1 and p_2 would be inconveniently large in dynes per sq. cm., other units may be employed. If g be omitted, p_1 and p_2 will be in gms. wt. per sq. cm. and

$$p_2 - p_1 = h\rho$$

This formula may also be used to calculate the pressure in metric tons (1,000,000 gms.) per sq. m. (10,000 sq. cm.) if h be in meters (100 cm.).

When British units are employed, the weight of a cylinder of 1 sq. ft. cross-section and h feet in length and of density ρ (lbs. per cu. ft.) is $h\rho$ lbs. Hence, if p_1 and p_2 are in lbs. wt. per sq. ft.,

$$p_2 - p_1 = h\rho$$

where ρ = specific gravity $\times$ 62.4 (§162).

188. Surface of Contact of Two Fluids.—The surface of contact of two fluids of different densities that are at rest and do not mix is horizontal. This may be deduced from the principle that, for stable equilibrium, the potential energy of a system must be a minimum (§113). If any part of the denser fluid were at a higher

191. Archimedes' Principle.—When a body is partly or wholly immersed in a fluid at rest, every part of the surface in contact with the fluid is pressed on by the latter, the pressure being greater on the parts more deeply immersed. The resultant of all these forces of pressure is an upward force, called the *buoyancy* of the body immersed. The direct calculation of this resultant force is difficult, except when the body has some simple form, such as a cylinder with its axis vertical; but a simple line of reasoning will give the magnitude and direction of the force.

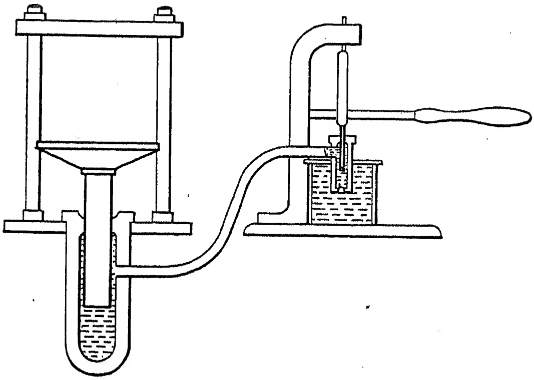


FIG. 88.—Hydraulic press.

The pressure on each part of the surface of the body is evidently independent of the material of which the body consists. So let us suppose the body, or as much of it as is immersed, to be replaced by fluid like the surrounding mass. This fluid will experience the pressures that acted on the immersed body, and this fluid will be at rest; hence the resultant upward force on it will equal its weight and will act vertically upward through its center of gravity. It follows that **a body wholly or partly immersed in a fluid is buoyed up with a force equal to the weight of the amount of the fluid that the body displaces.** The force acts vertically upward through the center of gravity of the fluid before its displacement, and the corresponding point in the immersed body is called its *center of buoyancy*.

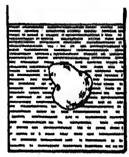


FIG. 89.

The weight, in dynes, of the fluid displaced equals the product of its volume (which equals the volume immersed), its density and g. Hence in c.g.s. units,

$$Buoyancy,\ B, = Volume\ immersed \times density\ of\ fluid \times g$$

If the g be omitted, the buoyancy will be in gms. wt. If g be omitted and V be in ft.^{3} and density be stated in lbs. per ft.^{3}, B will be in lbs.

Buoyancy is to be treated as any other force that acts on a body; it either causes motion or helps to produce equilibrium. If a body of weight w, wholly or partly immersed in a fluid, be sustained, partly by buoyancy B and partly by another upward force F, then,

$$F + B = w$$

192. Fluids in Motion.—While the calculation of the motion of a rigid solid body is comparatively simple, owing to the fact that we may treat a solid as a whole, without regard to the actions between its parts, the discussion of the motion of a fluid is rendered difficult by the readiness with which any part of the fluid changes its shape, and we cannot, therefore, without the use of advanced mathematics, treat of any except a very few and simple cases of the motion of fluids.

When a fluid moves, either in an open stream or in a closed pipe, the continual change of shape of each part is opposed by internal friction between these parts, and, to maintain the motion, some external force must be applied to the fluid. The most common causes of motions of fluids are gravity, as in the case of a river, pressure applied to some part of the boundary of the fluid, as in the case of water pumped through a system of piping, and the motion of solids in contact with the fluid, as in the case of a fan.

193. Flow in Pipes.—When the pressure on a fluid in a horizontal tube is greater at one end than at the other, a flow ensues. When the pressure is first applied, the motion begins with an acceleration, but in a short time, if the pressure at the ends is constant and the supply of fluid is maintained, a steady state of motion ensues, so that at each part of the tube the motion is constant. The simplest case is when the tube is of constant cross-section and the fluid is a practically incompressible one, that is, a liquid. In this case the volume of fluid passing all cross-sections of the pipe is the same throughout, and the rate of flow is, therefore, the same at each cross-section. The motion is from places of higher pressure to places of lower pressure. If, however, the fluid be compressible,

while the motion at a point remains steady and the mass that passes every cross-section is necessarily the same as that which enters the pipe, the volume of flow is variable; for where the pressure is greater the fluid is compressed into a smaller volume, and where the pressure is less the fluid is not so much compressed. Thus the speed of the fluid particles is, on the whole, greater in the parts of the pipe where the pressure is less, that is, the further along the stream we go.

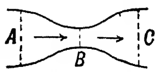


Fig. 90.

When a liquid flows through a tube of variable section (Fig. 90), the pressure at the ends being constant, the mass that passes each cross-section is the same, but the rate of motion of the particles increases as the stream comes to a contraction of the tube, and decreases again as the stream comes to an expansion of the tube. Now an increase of velocity, or an acceleration, necessarily means a smaller pressure ahead than behind, and a decrease of velocity necessarily means a larger pressure ahead than behind. Thus in a contraction (or "throat") the pressure is smaller than immediately before or behind, the amount of difference being dependent on the rate of flow through the tube and the cross-section at the throat and at either side. The general principle, of which this is an illustration, is that, along any streamline, that is a line of flow, in a fluid, *the pressure decreases wherever the speed increases.* This principle is the basis of the *Venturi meter* for gauging the flow of water in pipes. The same principle is employed in the *aspirator* (Fig. 91), a form of air-pump attached to a water faucet. The water is forced through a contraction in a tube and produces suction in a side tube which is connected to the vessel to be exhausted. Similar considerations apply to the flow of gas through a pipe of variable cross-section, but this case is complicated by the changes of density due to changes of pressure.

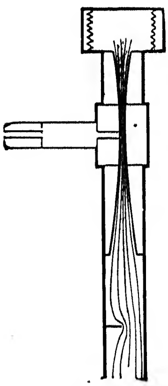


Fig. 91.

194. Illustrations of Law of Flow.—Fig. 92 represents (in section) a glass tube that passes tightly through a wide cork, and a second cork, through the center of which a pin is stuck. When air is blown through the tube, the lower cork is not repelled but is attracted (the pin prevents side motion). The air increases its speed to pass through the small space between the corks, hence its pressure diminishes, and atmospheric pressure pushes the corks together. Various other pieces of apparatus, such as the ball nozzle and the injector, act on the same principle.

The curvature of the path of a rotating base ball or tennis ball is due to a difference of pressure on the opposite sides of the ball. Suppose the ball had no translatory motion but had a motion of rotation, while a current of air blew on it, as indicated in figure 93. (This comes to the same thing since it is wholly a question of relative motion.) The rotating ball would carry air around with it. At A the two air motions would be in the same direction, and at B they would be in opposition. Hence the velocity of the air-particles would be greater at A than in the stream of air that blows on the ball, and at B the velocity would be less than in the general stream. The pressure at B would therefore be greater than that at A, the result being a force on the ball in the direction

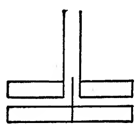


Fig. 92.

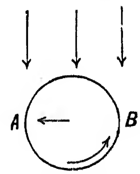


Fig. 93.—"Curve" of a base ball.

BA. The same result follows when the ball is moving through air otherwise at rest, hence the path curves toward the side of less pressure. The curvature of the path can also be explained by considering that the motion of the air carried around by the ball causes an increase of density, and therefore of pressure, at B, and a decrease of density and pressure at A.

195. Work Done by a Piston.—When a piston of area a moves a small distance d along a cylinder against a pressure P (per unit area), it exerts a force Pa through a distance d, and therefore does work Pad. Since ad is the volume, say ΔV, of the small

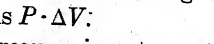


Fig. 94.—Work by a piston.

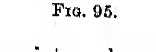


Fig. 95.

part of the cylinder through which the piston has moved, the work done is $P \cdot \Delta V$.

If the pressure is not constant, as, for example, when a gas in a closed vessel is being compressed, the whole work will be the sum of products $P \cdot \Delta V$, where P must be given its proper value for each successive change of volume ΔV. We may also use a graphical method as in §22 and §59. In the present case (Fig. 95), each abscissa will represent the volume at some moment, and the

corresponding ordinate will represent the pressure at that moment. The area will represent the whole work.

Conversely, an expanding fluid does work equal to the sum of $P \cdot \Delta V$, where P is the pressure and ΔV an increment of volume. If P is constant and the whole change of volume is V, the work done is PV.

196. Viscosity.—A fluid offers no *permanent* resistance to forces tending to change its shape; it yields steadily to the smallest deforming force. But the rate of yielding is different for different fluids, that is, different fluids offer different *transient* resistances to deformation. This transient resistance is called internal friction or *viscosity*. Thus a very viscous liquid, such as glycerin or tar, flows much more slowly through a tube or down an incline than water does, and such a flow consists in a continuous changing of shape of each part of the liquid. The internal forces are what we have called stresses, and, since the strain is a change of shape only, the stress must be a shearing stress.

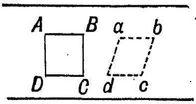


Fig. 96.—Shearing of a fluid.

Consider, as an example, a stream (Fig. 96) flowing down a very gentle incline under the force of gravity, and suppose the motion to be so slow and smooth that no eddies are produced. The speed is greater near the surface than at the bottom. A small cube $ABCD$ with sides vertical and horizontal will, by the motion, be changed into the form $abcd$. The liquid above AB exerts a force in the direction AB, on the upper face of the cube, and the liquid below CD exerts a resisting force on the face CD in the direction CD. These two forces constitute a shearing stress. Very extensive experiments have shown that **the ratio of the shearing stress to the rate of shear is a constant for any one fluid,** the value of the constant being different for different fluids. *The constant ratio of the shearing stress in a fluid to its rate of shearing is called the coefficient of viscosity of that fluid.*

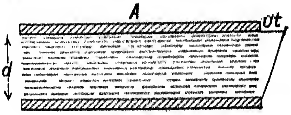


Fig. 97.

A concrete case will make this definition clearer and will lead to another way of stating the definition. Suppose (Fig. 97) that the space between two large parallel plates is filled with the fluid under consideration, and let one plate be moving slowly parallel to the other with a velocity v. Experiment (as stated later) shows that the fluid in contact with the plates does not slip along the faces of the plates but

adheres to them. The moving plate will, in a very short time t, move a distance vt, and, if the distance between the plates be d, the shear produced in the time t will be vt/d. Hence the *rate of shearing* is v/d. If the area of each plate be A and the force applied to move the upper plate be F, *the shearing stress* will be F/A. Hence, denoting the coefficient of viscosity by η, we have

$$\eta = \frac{F/A}{v/d}$$

$$F = \eta\frac{Av}{d}$$

If, now, we suppose A, v and d to be each unity, F will be equal to η. Hence we have the following way of stating the definition of η: The *coefficient of viscosity of a fluid is the tangential force on unit of area of either of two horizontal planes at unit distance apart, one of which is fixed, while the other moves with unit velocity, the space between them being filled with the viscous material.*

197. Measurement of Coefficient of Viscosity.—The most common method of finding η is by measuring the flow of the fluid through a tube of very small bore (or so-called capillary tube). The motion of the fluid in such a case (provided the velocity does not exceed a certain magnitude) is analogous to the slipping of the tubes of a small pocket telescope through one another. If we imagine the fluid divided into a very large number of thin cylindrical shells, the process consists in the slipping of shell through shell; hence the resistance encountered is internal friction or viscosity. Let p be the difference of pressure at the two ends of the tube (supposed horizontal), l the length of the tube, and r its radius. It has been shown theoretically and experimentally that, when the fluid is a liquid, the volume that flows out of the tube in unit time is

$$V = \frac{p\pi r^4}{8l\eta}$$

This is called *Poiseuille's law*. In deriving it the assumption is made that no slipping of the fluid along the surface of the tube takes place, and the agreement of theory and experiment confirms this assumption. The law also applies to a gas provided that p is small, but if p is large the formula must be modified to allow for the compressibility of the gas.

The following are some values of η in c.g.s. units at 20°C.

Alcohol	0.0011	Water	0.010
Ether	0.0026	Glycerin	8.0

The unit of viscosity is called the *poise*, *e.g.* the viscosity of water is 0.010 poise.

198. The Explanation of Viscosity.—Viscous resistance to fluid motion resembles friction between solids in certain respects, and in other respects the two are very different. Both are forces that appear only as resistances to relative motion: they are, therefore, non-conservative forces, and energy spent in doing work against them is changed into heat. But, while the friction between solids, is, through a considerable range of velocity, independent of the velocity, the resistance due to viscosity is exactly proportional to velocity through the widest range in which experimental tests have been made. This points to a fundamental difference in the nature of the resistance in the two cases.

The particles of a fluid are in motion and are not, like the particles of solids, confined to more or less definite positions. If, now, we imagine two layers of a fluid in relative motion, so that one is passing the other, like one railway train passing a second, it is evident that particles from each layer must be continually crossing the boundary into the other layer. The particles of the more rapidly moving layer that cross the boundary carry their larger momentum with them, and thus produce a gradual increase of the velocity of the second layer. At the same time particles of the latter layer penetrate into the former, and, by taking up momentum, diminish the velocity of that layer. The result, on the whole, is a tendency of the two layers to come to the same velocity, and this is exactly what we mean by a resistance to relative motion. In the case of gases, this explanation may be regarded as fully established: for the formulae to which it leads by mathematical methods are verified by experiment. In the case of liquids the process is probably complicated by something like interlocking of molecules, especially when the molecules of the substance are complex.

PROPERTIES OF LIQUIDS

199. Compressibility of Liquids.—While the shear modulus of any liquid is zero, the bulk modulus is usually large, that is, the pressure on a liquid must be greatly increased to produce much diminution of volume. The coefficient of compressibility of a liquid (§169) is therefore small. Measurements of the compressibilities of liquids are made by subjecting the liquids to great pressures in a vessel called a piezometer and noting the resulting diminution of volume.

The following table gives the compressibilities of some liquids. Each number is the fraction by which the volume of the liquid is decreased when the pressure on it is increased by one atmosphere.

Alcohol	0.0000828	Mercury	0.0000038
Ether	0.0001156	Water	0.0000489

200. Hydrometers.—A hydrometer is an instrument for finding the density of liquids; some hydrometers may also be used to find the density of solids. The action of most hydrometers depends on Archimedes' principle. Some hydrometers sink to different depths in different liquids and thus indicate the densities of the liquids; these are called hydrometers of *variable immersion.* Other hydrometers are used with different weights added to the weight of the instrument, so that they are always immersed to the same depth; these are called hydrometers of *constant immersion.*

The common hydrometer is one of variable immersion. It is a glass tube, with an enlargement in the middle, and weighted at the

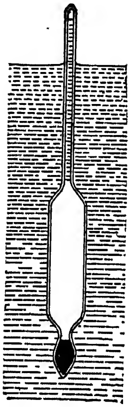


FIG. 98.—Common hydrometer.

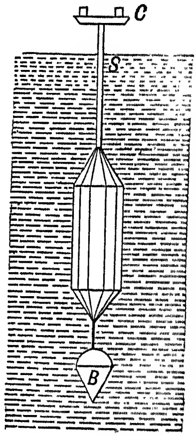


FIG. 99.—Nicholson's hydrometer.

lower end with mercury, so that it will float in stable equilibrium. Inside or on the tube is a scale which indicates the density of the liquid by the depth to which the tube is immersed (Fig. 98).

The best known hydrometer of constant immersion (Fig. 99) is Nicholson's hydrometer. It consists of a hollow cylindrical body (of metal or glass), to one end of which a somewhat heavy basket B is attached, while at the other end there is a stem S, which carries a scale pan C for weights. On the stem there is a mark indicating the depth to which the hydrometer is to be immersed. Let W be the weight of the hydrometer, and let w be the weight that must be placed on the pan C to make the instrument sink to the mark in water of density ρ, and w' the weight on C required when the hydrometer is in a liquid of density ρ'. The

volume of liquid displaced in both cases is the same. Hence, by Archimedes' principle, the weights of equal volumes of the second liquid and of water are $W + w'$ and $W + w$, and their ratio equals the ratio of the densities.

This hydrometer may also be used to find the density of a small solid. When so used the instrument is, in reality, a balance for weighing the solid in air and then in some liquid of known density. The body is first placed on C. The weight required on C to sink the hydrometer to the mark on the stem will be less than w', by the weight of the body. This gives the weight of the body in air. The body is then placed in B, and its apparent weight when immersed is found in the same way. The ratio of the weight of the body to its apparent loss of weight when immersed, which equals the weight of an equal volume of liquid, gives the ratio of the densities of solid and liquid.

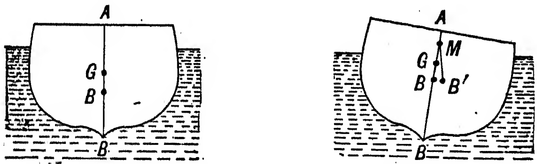


FIG. 100.—Stable equilibrium of a vessel.

201. Stability of Flotation.—A body floating at rest on the surface of a liquid is in equilibrium under the action of its weight, acting vertically downward through the center of gravity, G, and the resultant upward pressure of the liquid, acting through the center of buoyancy, B. Hence the two forces are equal and act in opposite directions in the line BG (Fig. 100). Suppose the body to rotate slightly about an axis perpendicular to the plane represented in the figure. The form of the volume of water displaced is now different (unless the body be spherical or cylindrical), and the center of buoyancy is at some point B' not in the vertical line through G. Hence the forces now acting on the body constitute a couple, and, if the couple tends to right the body, the equilibrium is stable; if not, it is unstable.

The simplest case to consider is when the body is symmetrical, or very nearly so, on opposite sides of the plane through B and G perpendicular to the axis of rotation; for in this case B' is in this plane. Let a vertical line through B' cut BG in M. For small rotations the position of M on BG is very nearly independent of

the magnitude of the rotation. M is called the *metacenter* of the body. If M is above G, it is evident that the couple tends to right the body, and the equilibrium is stable; if M is below G, the couple tends to displace the body further, and the equilibrium is unstable. This explains the danger of taking the whole cargo out of a vessel without putting in ballast and the risk of upsetting when several people stand up at the same time in a small boat. The metacentric height, GM, of a large steamship is usually only a foot or two, sometimes less than a foot.

202. Outflow from an Orifice—Torricelli's Theorem.—When an orifice is opened in a side of a vessel containing liquid, the liquid is forced outward. The simplest way of finding the velocity of the escaping liquid is by an application of the principle of the conservation of energy.

A small mass m of liquid escaping with velocity v has $\frac{1}{2}mv^2$ units of kinetic energy. If no liquid has been added to the vessel during the escape of the mass m, the potential energy of the liquid in the vessel must have diminished by an amount equal to $\frac{1}{2}mv^2$. The mass m was really removed from the part of the liquid near the orifice, but the change of the state of the liquid in the vessel is the same as if the mass m had been removed from the surface; and the change of total potential energy of the liquid in the vessel and of the escaping liquid is the same as if a mass m had been lowered from the surface to the depth of the orifice. Hence, if we denote the depth of the orifice below the surface by h, the loss of potential energy is mgh, and therefore

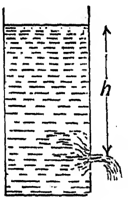


FIG. 101.

$$\tfrac{1}{2}mv^2 = mgh$$

and

$$v = \sqrt{2gh}$$

Thus *the velocity of escape is the same as if the escaping liquid had fallen freely through the distance of the orifice below the surface.* This is known as *Torricelli's theorem.*

The theorem relates only to the velocity of the particles as they leave the orifice. It does not enable us at once to calculate the volume that escapes in a given time: for the cross-section of the jet contracts for a short distance after leaving the vessel, and at a certain point reaches a minimum, called the *vena contracta* (or contracted vein), beyond

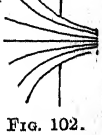


FIG. 102.

which it expands. If the area a of the cross-section of the *vena contracta* is found, the volume per second that escapes is av. The ratio of a to the area of the orifice depends on the velocity of escape and can be changed by inserting a tube (or *ajutage*) through the orifice.

203. Energy of a Stream. Bernoulli's Theorem.—When a liquid is flowing steadily, the path followed by a small part of the liquid is called a *stream line*. A tubular surface bounded by stream lines is called a *tube of flow*. So long as the motion is steady (§193), the stream lines and tubes of flow do not change in form or position. Through each tube of flow the liquid moves as through a pipe.

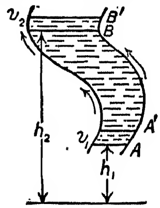


FIG. 103.

Consider the liquid that at any time occupies the part AA' of a tube (Fig. 103). A moment later it will fill the part BB', having been pushed forward by the liquid behind against the opposition of the liquid in front. The volume, V, of the part of the tube between A and A' is equal to that between B and B'. If the pressures at A and B are p_1 and p_2 respectively, the net work done is $(p_1 - p_2)V$ (§195).

Since the flow is steady, the kinetic energy in the part of the tube between A' and B is constant. Hence the net gain of kinetic energy is the same as if the part of the liquid of volume V and mass $V\rho$ were transferred from A to B and its velocity changed from v_1, the velocity of the stream at A, to v_2, the velocity at B. Thus the gain of kinetic energy is $\frac{1}{2}V\rho(v_2^2 - v_1^2)$. For a similar reason the gain of potential energy is $V\rho g(h_2 - h_1)$.

If we now equate work to increase of energy, we get

$$p_1 + g\rho h_1 + \tfrac{1}{2}\rho v_1^2 = p_2 + g\rho h_2 + \tfrac{1}{2}\rho v_2^2 = \text{constant}$$

This is called *Bernoulli's theorem*. It is of fundamental importance in hydraulics. From it Torricelli's theorem and the principle stated in §193 can readily be derived.

The equation, as written, may be used with either c.g.s. units or B.e. units. With the former p means dynes per cm,^{2} and ρ is in gms per cm.^{3} With B.e. units p means lbs. per ft.^{2} and ρ dynamical units of mass per ft.^{3} But in using the equation engineers usually state the density in lbs. per ft.^{3} and denote it by d. It is then necessary to modify the equation by replacing ρ by d/g.

204. Pressure Exerted by a Stream.—When a stream of liquid meets an obstacle and is arrested, it gives up its momentum to the obstacle, that is, it exerts a force on the obstacle. The pressure thus produced can be calculated from the velocity of the liquid and the amount of liquid that impinges per second on this obstacle. On this is founded a method of measuring the velocity of a stream (Pitot's tube). A tube, bent at right angles, is placed in the stream,

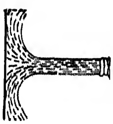


FIG. 104.

so that one arm points horizontally up stream and the other vertically upward. If the water were at rest, the liquid would rise in the vertical arm to the height of the surface of the water, but the pressure exerted by the stream raises it higher, and the velocity of the stream can be deduced from this additional height.

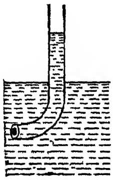


FIG. 105.—Pitot tube for finding velocity of stream.

In Pitot's tube the case of §202 is reversed. Let the rise of level be h. Suppose the liquid to be continually removed at the level h. In any time the total decrease of kinetic energy will be $\frac{1}{2}mv^2$ and the total increase of potential energy mgh. Hence $v^2 = 2gh$ (approx.), but a correction factor (slightly greater than unity) is necessary, because the tube disturbs the uniform flow of the stream.

When a jet impinges on an obstacle and flows off laterally, the pressure exerted is that due to the loss of the momentum of the liquid. If this obstacle is curved, so that the motion of the liquid is reversed, the water is given a momentum in the opposite direction, and the force exerted on the obstacle is increased. This principle is taken advantage of in the construction of water-wheels (Fig. 106).

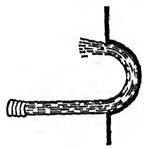


FIG. 106.

When a stream strikes an obstacle obliquely, it is partly arrested and then flows down along the surface of the obstacle. Thus the side of the obstacle farther up stream receives more momentum than the lower side, and so tends to turn more nearly perpendicular to the stream. A floating log, anchored at its middle point, sets itself across the stream. A leaf, falling from a tree, tends to take a horizontal position. The effect is readily illustrated by sweeping through the air a square disk of cardboard, connected by short threads to a wire frame (Fig. 107).

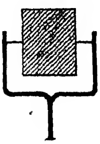


FIG. 107.—A disk swept through the air turns perpendicular to the direction of motion.

205. The Hydraulic Ram.—Water flowing under the action of gravity tends to the condition in which it would be in equilibrium, and in which, therefore, all parts of the free surface would be at the same level. This is the meaning of the statement that "water seeks its own level." Usually it is only by means of work done by some force other than gravity that water can be raised to a higher level. In the hydraulic ram a small fraction of the water in a stream is raised to a high level by a

self-acting mechanism that does not need any external power (Fig. 108).

When a stream of water in a pipe is suddenly stopped, for example, when a faucet is turned off, the momentum of the water, which may be very large, is stopped in a very short time, and therefore the force exerted by the water on the pipes may be very much larger than that which the water exerts after it has come to rest. In the hydraulic ram this momentary intense pressure is used to drive water into an air-chamber, such as is used in a force-pump.

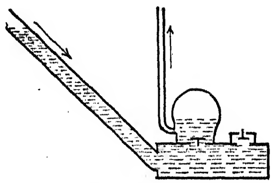


FIG. 108.—Principle of hydraulic ram.

Momentary interruptions of the current are caused by the opening and closing of a valve, which works automatically in a vertical direction. The weight of the valve is such that, when it is closed and the water is at rest, the pressure of the water on the lower surface of the valve is not sufficient to keep it closed; hence it opens and allows the stream to start. The stream, when in motion, carries the valve with it, again closing it and arresting the motion.

Some of the potential energy of the head of water is transformed into kinetic energy of the flowing stream, and this is partly changed into potential energy of the compressed air, which again is changed into potential energy of the water at the top of the delivery tube. Only a small part of the water is finally raised to a higher level than its original one, and its gain of potential energy is compensated by the loss of potential energy of the remainder.

MOLECULAR FORCES IN LIQUIDS

206. Molecular Forces.—Between the particles of a solid or of a liquid there are attractions that keep the body together, unless these forces are overcome by external forces. To show directly the existence of these forces between the particles of a liquid is very difficult, since a liquid changes its shape so readily. It has, however, been found possible to fill a glass tube with water at a high temperature and then seal the tube, so that the water, on cooling, continued to fill the tube, without contracting, until it exerted a pull of over seventy pounds per square inch upon the walls of the tube. The water would, in such an experiment, stand a much higher stress, if it were possible to free it perfectly from

absorbed gases. It is this attraction between the particles of a liquid that has to be overcome when a liquid is evaporated; and, from the heat required for evaporation, it can be calculated that the attractions between the particles are very powerful and produce a very great internal pressure across any imaginary plane in the liquid. This intrinsic pressure, as it is called, can also be calculated from the extremely small expansion when the pressure is greatly decreased. In the case of water it is about 10,000 times greater than atmospheric pressure.

From the above it might be thought that a body immersed in a liquid would feel the effect of this great internal pressure. That such is not the case is due to the fact that the molecular forces of attraction are negligibly small unless the distances between the particles are exceedingly small (§160). Now the thinnest solid that it is possible to insert in a liquid separates the particles so far that the attractions between them are negligible, and thus the pressure on an immersed solid is merely that due to the causes, gravitation and pressure on the boundary, considered earlier.

The distance to which the force of attraction is sensible is called the *range of molecular forces.* Any particle of a liquid is attracted by all particles that lie within this range, and these may be considered as contained within a sphere. This sphere, whose radius is the range of molecular forces, may be called the *sphere of influence* of a particle.

207. Surface Tension.—The molecular forces of which we have been speaking produce certain remarkable effects at the surface of a liquid. *The surface of a liquid tends to contract to the smallest area admissible.* Thus a drop of water falling through the air becomes spherical, since the sphere is the figure of least surface for a given volume. The same is true of a drop of liquid lead falling in a shot tower; the drop solidifies as it falls, and it is found to be spherical, when the fall is sufficient to allow it to become perfectly solid while in the air. A mixture of alcohol and water can be prepared of the same density as an oil, and a large drop of the mixture, floating totally immersed in the oil, is spherical. When the end of a stick of sealing-wax or of a glass rod is melted in a flame, it tends to the spherical form.

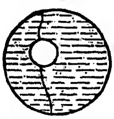


Fig. 109.—Loop of silk on surface of film.

A beautiful illustration of the tendency of a liquid surface to contract consists in forming a film from a soap-bubble solution on a ring

of wire, to which a loop of silk has been loosely attached, so that the loop floats in the film; when the film is broken inside the loop the latter becomes circular. In shrinking to the form of least area the film pulls the loop into the form of greatest area for a given periphery, and this is a circle (Fig. 109).

208. Explanation of Surface Tension.—Consider the condition of a particle at A in the body of a liquid, and that of a particle at B, at less than the range of molecular forces from the surface. The particle at A is equally attracted on all sides by the particles around it, but the particle at B is more attracted inward than outward, since a sphere with center at B and the range of molecular forces as radius lies partly outside of the liquid. To take a particle from A to B, work must be done against this inward attraction.

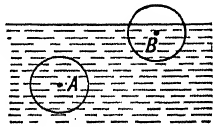


FIG. 110.

Now, when the surface of a liquid is increased, for example when a soap film is stretched, more particles are drawn into the surface; hence some work is done by the stretching force, and therefore an opposing force is overcome. But the stretching force required is parallel to the surface; hence the liquid exerts an opposing or contractile force parallel to the plane of the surface, and this force is what we call the surface tension. Thus the existence of a tension in the surface of a liquid is consistent with the principle of work. But this does not tell us how the surface molecules coöperate in resisting a stretching of the surface. To get at this has been difficult. Recent work seems to show that molecules are not spherical but more or less elongated or flattened. In the body of the liquid they form groups with their axes in the same direction, the direction being different for different groups, but in the surface, owing to the forces between the electric charges which they contain, they line up with their axes parallel. This seems promising, but much more remains to be learned.

FIG. 111.—Surface tension is the force across unit length.

209. Methods of Measuring Surface Tension. If a line be imagined drawn along the surface of a liquid (Fig. 111), the part of the surface on one side of the line pulls on the part on the other side, and if the length of line be supposed one cm. the pull in dynes is taken as the magnitude of the surface tension T, of the liquid. *The magnitude of surface tension is the force of contraction across a line of unit length in the surface.* Almost any of the effects

of surface tension may be made the basis of a method of measuring it. When the liquid can be formed into a thin sheet, as in the case of a soap solution, a direct method of measuring it may be used (Fig. 112); a film may be formed on a wire frame, of which one side is movable; if the force required to hold this side at rest against the surface tension is F, and the length of the movable side is l, the tension in each surface of the film is $F/2l$.

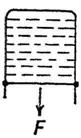


FIG. 112.—Stretching a film.

To draw a horizontal wire up through the surface of a liquid the tension of the surface must be overcome, and the surface tension can be calculated from the force required.

The movement of minute waves or ripples on the surface of a liquid is due chiefly to the surface tension of the liquid, and, from the wave-lengths of the ripples and their velocities, we can find the magnitude of the surface tension.

The rise of a liquid in a capillary tube depends, as we shall see later, on the surface tension of the liquid, and this affords another method of measurement.

The following table gives (in dynes per cm.) the surface tension between some liquids and air (at 20°C.)

Alcohol	24.5	Olive oil	34.3
Ether	17.6	Mercury	539
Water	73.5		

Surface tension decreases with rise of temperature. For water the decrease is about 0.14 per degree C., for alcohol about 0.09.

210. Contact of Liquid and Solid.—The general free surface of a liquid is horizontal; but, where the liquid is in contact with a solid, the surface is usually curved, the direction and amount of the curvature being different for different liquids and different solids. Water in contact with a vertical surface of glass is curved upward, and mercury in the same circumstances is curved downward. These, for a reason stated later, are called *capillary* phenomena.

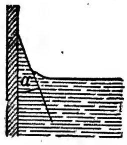


FIG. 113.—Contact of water and glass.

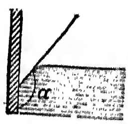


FIG. 114.—Contact of mercury and glass.

The contact angle of the wedge-shaped part of the liquid, between the free surface of the liquid and the surface of the solid, is called the *angle of capillarity*. The size of the angle, in any case, depends on the purity of the liquid and the cleanness of the solid surface. Thus, for very pure water in contact with clean glass, the angle is 0°; but, with slight contamination, even such as is caused by exposure to air, the angle may become as large as 25° or more. For perfectly pure mercury and glass the angle is about 148°, but slight contamination reduces it to 140° or less; for turpentine it is about 17°, for petroleum 26° and so on.

211. Level of Liquids in Capillary Tubes.—When a glass tube of very fine bore (or so-called capillary tube), open at both ends, is placed vertically, with its lower end in a vessel of liquid, the surface of the liquid in the tube is usually higher or lower than the general surface in the vessel. When the liquid is water or alcohol, the surface is elevated in the tube; when the liquid is mercury, the surface is depressed. For a given liquid the amount of elevation or depression is greater the smaller the bore of the tube, being, in fact, inversely as the diameter of the bore. For tubes of other materials than glass similar effects, depending in amount on the material of the tube, are observed.

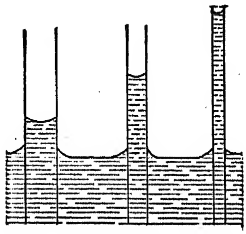


FIG. 115.—Water in capillary tubes.

There are similar elevations and depressions between two glass plates standing close together in a liquid. The elevations and depressions and the curvature of a liquid surface in contact with a solid are usually grouped under the general title of *Capillarity*.

Assuming the existence of the invariable angle of capillarity at which a liquid meets a solid, we can give a simple explanation of capillary elevations and depressions.

Consider the case when the liquid is elevated (Fig. 116). The liquid in the tube meets the tube in a circle of radius r, equal to the radius of the bore, and, at every point of the circle, the angle of contact is the angle of capillarity α. Thus the surface tension of the liquid pulls on the tube in the direction PQ inclined at α to the length of the tube; and the tube therefore reacts with an equal pull in the direction QP. The amount of the pull per unit length of the circumference of the circle of contact is T, and the component of this, taken parallel to the length of the tube, is $T.\cos\alpha$. For the

whole circumference of the circle of contact the sum of these components is $2\pi rT \cos \alpha$. This is an upward force on the liquid in the tube, and it draws the liquid upward, until the weight of the liquid elevated above the ordinary surface equals the supporting force. If the mean elevation is h cms., the volume of the supported column is $\pi r^2 h$ cm³ and its weight $\pi r^2 h\rho g$ dynes. Hence

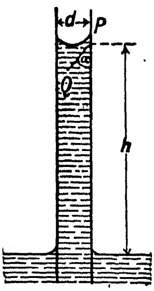


FIG. 116.

$$\pi r^2 h\rho g = 2\pi rT \cos \alpha$$

and

$$h = \frac{2T \cos \alpha}{g\rho r}$$

Thus the elevation is directly as the surface tension and inversely as the radius of the tube. By measuring the elevation and the radius and finding α by some other method, the value of T for any liquid may be obtained.

212. Elevation between Plates.—The above method of proof may also be extended to the case of a liquid between parallel plates (Fig. 116). In this case the surface of the liquid meets the surfaces of the plates in straight lines. Let the distance between the plates be d. Consider the equilibrium of the liquid contained between the plates and two vertical planes, perpendicular to the plates, and at unit distance apart. The pull of the surface tension at the top is $2T \cos \alpha$, and the weight of the liquid supported is $dh\rho g$. Hence

$$h = \frac{2T \cos \alpha}{g\rho d}$$

Thus the elevation is the same for two parallel plates as for a tube, if the distance between the plates equals the radius of the tube.

213. Pressure Caused by a Curved Surface under Tension.—Since the liquid in a capillary tube is elevated above or depressed below the ordinary level, the pressure beneath the curved surface must be less or greater than the pressure at the general surface. When the effect is a depression (mercury in glass), the depressed surface is curved downward, and the tension in the surface produces a pressure, just as the tension in a rubber sheet stretched over a ball produces pressure on the ball. When the effect is an elevation, the stretch on the upward curved surface tends to draw the liquid in the surface layer away from the liquid below, and so produces a state of tension or diminution of pressure beneath the surface.

From the amount of the elevation or depression we can calculate the change of pressure thus caused. In the case of an elevation to a height h, the pressure must be less than the pressure at the ordinary level, which is atmospheric pressure, by $g\rho h$, or (§211) $(2T\cos\alpha)/r$. Here r is the radius of the tube. If we denote the radius of the spherical surface by R, $R\cos\alpha = r$. Hence the pressure beneath the concave surface is less than that of the atmosphere above by $2T/R$. The same applies to the pressure produced on the concave side of a depressed surface. This difference of pressure on the two sides is due entirely to the tension and the curvature of the surface.

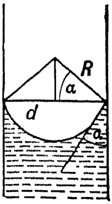


FIG. 117.

In the case of a spherical soap-bubble, there are two surface tensions to be considered, one on the inner side of the film and the other on the outer side. Hence the excess pressure inside the bubble, due to the tension and curvature of the film, is $4T/R$.

A cylindrical surface in a state of tension also produces pressure on the concave side. This is deduced, as above, from the elevation or depression of a liquid of surface tension T between two parallel plates at a distance d apart. If R is the radius of the cylindrical surface of the liquid, $R\cos\alpha = \frac{1}{2}d$. Hence (§212) $p = T/R$, and this is therefore the pressure on the concave side, due to the tension T in a cylindrical surface of radius R. In the case of a cylindrical soap-bubble of radius T the tension in each surface produces pressure T/R. Hence the pressure inside is greater than that outside by $2T/R$.

214. Other Effects of Surface Tension.—When the angle of capillarity of a liquid in contact with a solid is small, the liquid, in its attempt to establish this small angle, spreads out on the surface of the solid: that is, the liquid is one that *wets* the solid. Thus a drop of water, let fall on clean glass, spreads out, the angle of capillarity being small. A drop of mercury on a glass plate has no tendency to spread, but gathers into a ball.


FIG. 118.—Water between glass plates.

A film of water between two glass plates makes it difficult to draw the plates apart by a force normal to their surfaces. The liquid tends to spread over both plates and becomes concave outwards, so that the pressure within it is less than the atmospheric pressure on the outside of the plates, and this produces an apparent attraction between the plates.

When an attempt is made to blow out a glass tube containing numerous drops of water, a surprising resistance is experienced. Each drop becomes concave on the side of high pressure and the total resistance is the sum of the pressures exerted by these concave surfaces.

Small bodies, such as straws and sticks, floating on the surface of a liquid usually attract and gather into groups. Let us represent two such bodies by small vertical plates (Fig. 119). If the liquid wets both, it rises between them, and the pressure in the elevated portion is less than the atmospheric pressure on the outer side of the plates. Hence the plates are pushed together. If the liquid does not wet either plate, it is depressed between them; the pressure above the depressed part is atmospheric, while the pressure in the liquid on the outer sides of the plates is greater than atmospheric, and the plates are pushed together. If the liquid wets one plate but not the other, there is a part of each plate where the pressure on the inside is greater than that on the outside; hence an apparent repulsion results. (Pellets of paraffin wax, some of which are lamp-blacked, floating on water, will illustrate all three cases.)

Any dissolved substance or impurity changes the surface tension of water. This explains the irregular motions of small particles of camphor, dropped on clean water. At some points the camphor dissolves more rapidly than at other points, and near the former the surface tension of the water is weakened, so that the pull on the opposite side where the tension is greater, prevails and causes irregular motion.

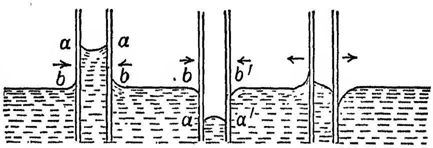


FIG. 119.—Capillary attractions and repulsions.

215. Diffusion of Liquids.—The gradual mixture of two liquids which come into contact is called *diffusion*. It takes place on a large scale where fresh water from a river flows out into the ocean. It may be illustrated on a small scale by pouring a solution of a colored salt into a tall vessel, and then cautiously covering the colored solution with a layer of water. The particles of each liquid are in motion and begin to make their way across the interface, and, after a long time, the whole vessel is filled with a mixture of the same constitution throughout. Stirring has the effect of increasing the area of contact of the liquids and so promotes diffusion. But some substances, such as oil and water, will not diffuse into one another completely, or "mix," probably because of the ways in which molecules are arranged in their surface films.

Let us denote the diffusing liquids by A and B, and let us suppose that initially A occupies the lower half of a tall jar and B the upper half. The *concentration* of either of the liquids at any point is its mass per unit volume at that point (*i.e.*, its density at the point, if the other liquid be imagined absent, without the first being

disturbed). The liquid A diffuses vertically upward, that is, from places of high concentration to places of low concentration. The *gradient of concentration* in any direction is the rate at which the concentration falls off in that direction; if the rate of fall per unit of distance is unity, the gradient of concentration is unity.

The general law of diffusion is that *the rate of diffusion for each liquid is proportional to the gradient of concentration* of that liquid. The *coefficient of diffusion,* or the *diffusivity* of the liquid, is the mass in grams that crosses unit area in a day, when the gradient of concentration is unity. This constant can be found from observations of the density at various points along the direction of diffusion, made by means of beads of different densities floating in the liquid, and in various other ways. The following table contains the coefficients of diffusion of various substances into water at the temperature (Centigrade) stated.

Hydrochloric acid	1.74 at 5°
Common salt	0.76 at 5°
Common salt	0.91 at 10°
Sugar	0.31 at 9°
Albumen	0.06 at 13°
Caramel	0.05 at 10°

From the above it will be seen that liquids vary widely in diffusivities. Substances of high diffusivity are called *crystalloids,* and those of low diffusivity are called *colloids.* The former group includes mineral acids, salts and substances generally that form crystals (whence the name), while the latter includes gums, albumen, starch, and glue (the name being derived from the Greek for glue). Crystalloids, dissolved in water, produce many marked changes in its properties; a colloid in water forms a jelly, which seems to consist of a semi-solid frame work holding the liquid in its meshes. Colloids have large and complex molecules, and it is, perhaps, to this fact and to the consequent slower motions of the molecules that their small diffusivities are due. They are comparatively tasteless, as they do not diffuse and reach the nerve terminals. Their low rates of diffusion also render them indigestible. Through a layer of a colloidal jelly crystalloids will diffuse almost as rapidly as through water, but colloids not at all.

216. Diffusion through Membranes. Osmosis.—Through certain membranes that have no visible pores, many liquids will diffuse readily. Thus, through a partition of rubber between water and

alcohol the alcohol will pass rapidly, while the passage of the water is barred. Water passes readily through animal membranes that it wets. A method of separating crystalloids and colloids, called *dialysis*, depends on the different rates at which these substances pass through such a membrane as parchment paper. The diffusion of substances through such septa is called *osmosis*.

Some membranes allow one constituent of a mixed liquid or solution to pass, while barring the other constituent: such membranes are called *semi-permeable*. One such is ferrocyanide of copper, formed in the pores of a porous partition by the reaction between ferrocyanide of potassium on one side and copper sulphate on the other. When such a membrane separates water and the aqueous solution of any one of various salts, the salt does not pass, but the water passes in both directions, though more rapidly toward the solution than in the opposite direction. If the solution be in a tube (Fig. 120) the lower end of which, closed by a plug of the membrane, is dipped in water, the level in the tube will rise, until (provided the membrane does not break) the column is of such a height that its pressure prevents further flow. This pressure is called the *osmotic pressure* of the solution. Its magnitude, for very weak solutions, is proportional to the concentration, that is, to the number of molecules of the dissolved salt per unit volume. For a large number of salts the pressure is the same for solutions that contain the same number of molecules of the salt in unit volume. For various other salts the osmotic pressure, for a given number of molecules per unit volume, is two (or some whole number of) times greater than for the first group; this is possibly due to the molecules being separated into atoms (or rather "ions") in the solution, the atoms acting independently. But the full explanation of osmosis and osmotic pressure is a matter of much dispute. One remarkable fact may be noted, namely, that the osmotic pressure for a given number of molecules (or of dissociated atoms) in an aqueous solution is equal to the pressure that these molecules (or atoms) would produce, if freely flying as gaseous particles in the space occupied by the solution. It is also noteworthy that the osmotic pressure increases in the same way and at the same rate with rise of temperature as the pressure of a gas does.

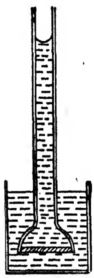


FIG. 120—Osmotic pressure.

Osmosis plays an important part in many processes that take place in the bodies of animals and plants, but the details belong to Biology.

PROPERTIES OF GASES

217. A gas has already been defined as a fluid that has no definite volume of its own, independent of the containing vessel, but expands so as to occupy any vessel in which it is contained. Gases have the same properties as liquids in all respects that depend on the fact that the shear modulus of a fluid is zero. The pressure at a point in a gas is the same in all directions (§184). The thrust of a gas on a surface is normal to the surface (§183). Pressure applied to any part of the boundary is equally transmitted in all directions (Pascal's principle §189). A body immersed in a gas is buoyed up with a force equal to the weight of the gas displaced (Archimedes' principle §191). The pressure in a gas increases with its depth at a rate expressed by $g\rho h$, as in the case of liquids (§185). Gases also show the property of internal friction or viscosity, and the definition of the coefficient of viscosity of a gas is the same as that of a liquid. Some of these properties are of special importance in the case of a gas and call for separate treatment.

218. Pressure of the Atmosphere.—A very important example of the pressure of a gas is the pressure exerted by the earth's atmosphere. The atmosphere, consisting chiefly of oxygen and nitrogen, is held to the earth by the gravitational attraction between it and the earth. The force that it exerts on any horizontal area of the earth's surface is the weight of all the air vertically above that area. At the top of a mountain the pressure is less than at sea level, since less of the atmosphere is above.

FIG. 121.—Torricelli's experiment.

Galileo discovered that air had weight by weighing a glass globe containing air, and then reweighing it when he had forced more air into it. His friend and pupil Torricelli found (in 1643) that, when a tube 33 inches long (Fig. 121), filled with mercury and closed at one end, was inverted in a dish of mercury, the mercury stood at a height of about 30 inches in the tube, thus leaving a vacuum above. This is known as *Torricelli's experiment.* From this he inferred that

the force exerted by the atmosphere on any area equals the weight of a column of mercury about 30 inches high and of a cross-section equal to the area. On hearing of Torricelli's experiment, Pascal reasoned that the pressure should be less and the column of mercury in Torricelli's tube shorter at the top of a mountain, and he wrote to a relative, who lived near the Puy de Dome in Auvergne, to make the test. The result confirmed his conjecture.

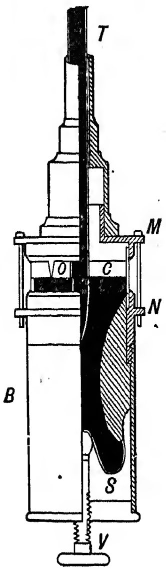


Fig. 122.—Cistern of Fortin's barometer.

219. The Mercurial Barometer.—Torricelli's tube was the first and simplest barometer, or pressure-gauge for measurement of the pressure of the atmosphere. The most accurate mercurial barometer of the present day is a Torricellian tube, with a scale and vernier for accurate measurement of the height of the mercury column, and a device by which the mercury in the cistern may be readily brought to a definite height. In Fortin's *cistern barometer* (Fig. 122) the cistern, C, has a leather bottom, S, the center of which rests on a screw, V. By turning the screw, the level of the mercury in the cistern can be raised or lowered, so that, when the barometer is read, the level of the mercury in the cistern shall always be the same, namely, zero of the scale on which the height of the barometer is read. Without such an adjustment, the level of the mercury in the cistern would fall or rise as the height of the mercury in the tube, T, rose or fell. That the level of the mercury in the cistern may be observed, the upper part of the cistern is of glass, and a small ivory stud, O, projecting downward from the top of the cistern, is adjusted by the maker, so that its end is on a level with zero of the scale. The image of the stud in the surface of the mercury is observed, and when, as the level of the mercury is raised by the screw, the end of the stud and the end of its image just meet, the surface of the mercury is at the zero of the scale. In filling such a barometer care must be taken that no air remains in the mercury, and, for this purpose, after the tube has been filled, the mercury is boiled, so that the air is expelled. The mercury in the cistern becomes somewhat tarnished in course of time, and the image of the stud ceases to be distinct.

A simpler form of barometer is Bunsen's *siphon barometer* (Fig. 123). In this there is no cistern, but the lower end of the tube is turned vertically upward. The difference of level in the open and in the closed end is the barometric height. Thus readings of both ends of the mercury column are necessary. Scales are etched on both branches; the one on the longer arm reads upward, and that on the shorter arm reads downward. The two scales are usually laid off with the same position for the zero, so that the sum of the two readings is the height of the barometer.

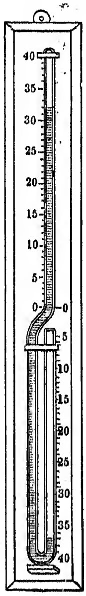


FIG. 123.—Bunsen's siphon barometer.

Another form of barometer is the *aneroid* (Greek anēros = dry) barometer in which no liquid is used. It consists of a metallic box, exhausted of air, with a thin metallic cover. Changes in atmospheric pressure cause slight changes of curvature in the cover, and by means of a multiplying system of levers these changes are transmitted to a pointer, which moves around a circular scale that is graduated in cms. or inches, so as to correspond to the readings of the mercurial barometer. This form of barometer is more convenient for travellers, but it has the disadvantage that its index must frequently be reset by comparison with the mercury barometer.

220. Uses of the Barometer.—A knowledge of barometric pressure is of great importance in weather forecasting. The governments of the United States and other civilized nations maintain a large number of stations, where records of the barometer are kept. From simultaneous readings over a wide area, the direction in which storms (or areas of low pressure) will move can be predicted. Such predictions lead annually to the saving of thousands of lives and of much valuable property in shipping.

Since the atmospheric pressure is less at higher levels, it is possible to ascertain the height of a mountain by observing the atmospheric pressure at the top and at the bottom. Near sea-level the height of the barometer diminishes by about 0.1 inch for every 80 feet of ascent; but as the elevation increases the rate of fall diminishes, owing to the greater rarity of the air. Allowance must be made for any difference of temperature at the two stations of observation. *Altimeters* for airplane use are aneroid

barometers with scales reading altitude. *Air-speed gauges* are also aneroid instruments, indicating speed through the air by the pressure (or suction, §193) produced in a tube.

221. Pressure and Volume of a Mass of Gas.—Common observation shows that added pressure on a mass of gas diminishes its volume. Thus, in pumping up a pneumatic tire, a large volume of air from the atmosphere is forced by high pressure into the small volume of the tire. Conversely, diminution of pressure allows a gas to expand. Against the pressure exerted on a gas it exerts an equal and opposite pressure, so that it is immaterial whether we speak of the pressure *on* or pressure *of* a gas.

FIG. 124.—Boyle's tube for pressures greater than atmospheric.

The law connecting the volume and the pressure of a gas is extremely simple, but it was not discovered until 1662, the discoverer being Robert Boyle. (Fourteen years later Mariotte rediscovered the same law.) *The volume of a gas at constant temperature varies inversely as its pressure,* or **the product of the pressure and the volume of a gas at constant temperature is constant.** Boyle discovered this law by experiments conducted with a tube bent as in Fig. 124, the shorter arm being closed and containing air and mercury, while the longer was open and was filled to varying depths with mercury. If, to the difference of level in the two arms, the height of the mercury barometer at the time is added, the sum is proportional to the pressure on the air, while the length of the tube occupied by air is proportional to the volume of the air. Thus he discovered the truth of the law for pressures exceeding an atmosphere. For pressures below an atmosphere he used a straight tube (Fig. 125) containing, initially, air and mercury and closed at one end; the open end was then plunged into a deep vessel of mercury. When the tube was drawn to different heights, the volume of the air increased with diminishing pressure. Thus Boyle verified the law for pressure less than an atmosphere.

FIG. 125.—Boyle's tube for pressure less than atmospheric.

222. Deviations from Boyle's Law.—While the law stated by Boyle is accurate enough for all ordinary practical purposes, careful tests have shown that it is not perfectly accurate. The most com-

plete tests were made by Amagat. He found that in the case of air, while the pressure is being increased from one atmosphere to about 78 atmospheres, PV steadily diminishes, until its value is 0.98 of its value at one atmosphere. Thereafter, with increasing pressure, PV increases. In the first stage (that is, up to 78 atmospheres) V decreases more rapidly than Boyle's law would indicate; thereafter, it decreases less rapidly, so that at 3000 atmospheres its volume is 4.2 times what it would be if Boyle's law were perfectly accurate. At 3000 atmospheres air has a density of 0.93, nearly equal to that of water.

Other gases show similar deviations from Boyle's law, but the pressure at which PV is a minimum is widely different for different gases, and so, too, is the magnitude of this minimum value of PV.

Starting with the view that a gas consists of flying particles, the impacts of which produce the pressure observed in a gas, van der Waals deduced the following formula, which agrees much better (though still not perfectly) with the results of experiments:

$$(P + a/V^2)(V - b) = \text{a constant},$$

at constant temperature, a and b being constants that are different for different gases. The term a/V^2 added to P takes account of the internal pressure caused by the mutual attractions of the molecules. By subtracting b from V allowance is made for the actual volumes of the molecules, as distinguished from the volume V of the space in which they dash about. From the value of b found by experiment the sizes of molecules have been deduced.

223. Modulus of Elasticity of a Gas.—The shear modulus of a gas being zero, a gas has only one modulus, namely the bulk modulus, and this is (when the gas is kept at constant temperature) simply equal to the pressure, P, of the gas. This is a consequence of Boyle's law. For when the pressure is P and the volume V, let an additional small pressure p be applied, and let the volume be thereby reduced by the small quantity v. Then by Boyle's Law

$$(P + p)(V - v) = PV$$

or, if we neglect the product of the small quantities p and v,

$$Vp = Pv$$

Now the bulk modulus is the increase of pressure, p, divided by

the proportional decrease of volume, v/V, and from the last equation this is equal to P.

By differentiating $PV = C$ we get $PdV + VdP = 0$ and from this we get $dP \div (-dV/V) = P$ The term $-dV$ means a decrement of V, as required in the definition of bulk modulus.

224. Buoyancy of a Gas.—A balloon containing hydrogen or helium, being lighter than the volume of air which it displaces, ascends in the air when released. The force giving it an acceleration upward equals the difference of its weight and the weight of the air that it displaces. If it rises to such a height that its mean density equals the density of the rarefied atmosphere, it will not ascend higher, unless lightened by casting some of its load overboard. The greatest height to which a manned balloon has risen, $13\frac{3}{4}$ miles, was reached by the *Explorer II* of the U. S. Army in 1935. Unmanned balloons have risen much higher. A large man displaces about $\frac{1}{4}$ lb. of air. When a body is weighed in air with weights that are supposed correct if used in a vacuum, to get the true weight correction must be made for the effect of the buoyancy of the air.

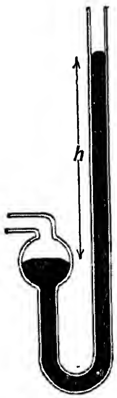


FIG 126—Open tube manometer

225. Manometers.—A manometer is an apparatus for measuring the pressure of a fluid. In the simplest form (Fig. 126) the pressure to be measured is balanced against the pressure of a column of liquid in a tube. This is called the *open tube manometer* or siphon gauge. The pressure is found from the difference of level of the liquid in the two arms and the density, ρ, of the liquid. When expressed in dynes per cm.^{2}, $P = h\rho g +$ atmospheric pressure, while in gms. per cm.^{2} or lbs. per ft.^{2}, $P = h\rho +$ atmospheric pressure.

In another manometer (Fig 127) the pressure to be measured is balanced against that of a gas (usually air) in a uniform *closed tube*. By Boyle's Law the pressure in the gas is inversely as the volume, that is, inversely as the length of the air column. The pressure in the gas plus that indicated by the difference of level of the liquid is the pressure to be measured

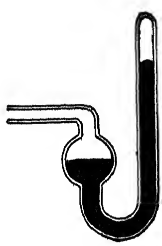


FIG. 127.—Closed tube manometer.

In *Bourdon's Pressure Gauge* (Fig 128) a hollow tube of metal having an elliptical cross-section is bent into an arc of over 180°.

One end of the tube is closed. When the fluid of which the pressure is to be measured is admitted to the open end, the curved tube becomes less curved under the increased pressure, and more curved under decreased pressure. An index moving over a scale is attached to the free end. The action depends on the fact that the pressure tends to increase the interior volume of the tube; and, since a circular cross-section allows of more volume than an elliptical one for a given periphery, the section will, under increased pressure, tend to the circular form, and the change of form of the cross-section causes the change of shape of the tube.

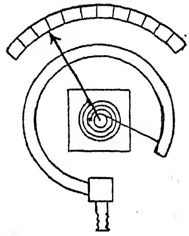


FIG. 128.—Bourdon's pressure gauge.

226. Viscosity of Gases.—The viscosities of gases are small compared with those of liquids. Thus the viscosity of air is about $\frac{1}{80}$ of that of water. While the viscosity of air is small, it is sufficient to retard greatly the fall of small particles of dust and small drops of water, such as constitute a cloud. In a cloud (where the air may be one thousand times less dense than water) a drop of water one thousandth of an inch in diameter falls about 0.8 inch per second, while a drop one ten-thousandth of an inch in diameter falls about one hundred times more slowly, or about 0.5 inch in a minute. For large drops, such as constitute rain, the viscosity of air offers practically no resistance; the resistance that prevents such drops attaining enormous velocities is the inertia of the air.

The viscosity of a gas increases when its temperature rises, which is the opposite of the case with liquids. The viscosity of a gas at constant temperature does not change appreciably when its density is considerably altered by change of pressure.

227. The Kinetic Theory of Gases.—The view that a gas consists of a myriad of particles in incessant motion may be regarded as firmly established. The evidence for this belief is that we can from it deduce nearly all the mechanical properties of a gas, and the general agreement between these deductions and the observed facts could hardly be a mere accidental coincidence. As we do not yet know all about the structure of the particles of which a gas consists, there are some properties that cannot yet be fully explained by this theory. A definite contradiction between the known properties and the deductions made from the theory

would be fatal to the latter, but no such contradiction has been found.

As an illustration of the way in which the theory accounts for the properties of gases we shall show that it explains Boyle's Law.

Before doing so we must state the theory more in detail. The following, while an incomplete statement, will be sufficient for our purposes. (*a*) A single gas consists of particles all of the same size moving in random directions; (*b*) when the particles impinge on one another and on the walls of the vessel, they rebound like perfectly elastic smooth spheres, with a coefficient of restitution of unity; (*c*) unless a gas is greatly condensed, the particles are so far apart, compared with their dimensions, that the forces they exert on one another may be neglected, except at impact We do not assume that the velocities of all the particles are the same, and, in fact, there is good ground for believing that the velocities differ considerably. Impacts between particles are probably more like sudden repulsions than actual contacts. If they caused a loss of kinetic energy, all the air molecules in a room would soon fall on the floor.

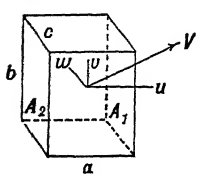


FIG. 129

By assuming that the particles are all similar we are confining ourselves, in this brief outline of the subject, to a gas that has no isotopes (§159), but it can be shown that the presence of isotopes does not affect the result.

For simplicity, consider a gas contained in a rectangular vessel, the edges of which are a, b and c in length, and let the faces A_1 and A_2, each of area bc, be perpendicular to the edges of length a. Let us first fix our attention on some particular particle which has a velocity V in some direction. V may be resolved into three components, u, v, w, in the directions of the edges respectively, u being in the direction of the a edges Suppose the particle to impinge on the side A_1. The impulse that it will give to that side at impact will depend on its mass and on u, not at all on v and w. If it impinged without rebounding, it would give to A_1 momentum equal to its mass, m, multiplied by u, or mu. But it rebounds with a velocity the component of which perpendicular to A_1 is u in the opposite direction: hence the momentum it gives to A_1 is $2mu$. Let us now suppose, for the present, that it reaches A_2 without impinging on any other particle; for this it will require a/u seconds. At A_2 it will rebound with a velocity the component of which perpendicular

to A_2 is u, and will, supposing it to encounter no other particle, reach A_1 in time $2a/u$, when it will again rebound. Hence in every second it will impinge $u/2a$ times on A_1, and in every second it will give to A_1 momentum $2mu \cdot u/2a$ or mu^2/a. The total force exerted on A_1, that is the momentum imparted to A_1 per second, is the sum of mu^2/a for all the particles, and to find the pressure p on A_1, we must divide this sum by the area of A_1, namely bc. Hence

$$p = \frac{m}{abc}(u_1^2 + u_2^2 + \ . \ . \ .)$$

Let us now denote the total number of particles in the vessel by N, and the number per unit volume by n. Since abc is the total volume of the vessel, $nabc = N$. Hence

$$p = mn\frac{(u_1^2 + u_2^2 + \ . \ . \ .)}{N}$$

The product mn is the mass of all the particles in unit volume, that is, the density ρ; and $(u_1^2 + u_2^2 + \ . \ . \ .)/N$ is the average value of u^2 for all the N particles in the vessel. Denoting this by $\overline{u^2}$ we see that $p = \rho\overline{u^2}$. For any one particle

$$V^2 = u^2 + v^2 + w^2,$$

and, since the particles are moving wholly at random, the average values of u^2, v^2 and w^2 are all equal, and the value of each is therefore $\frac{1}{3}$ of the average value of V^2, which we may denote by $\overline{V^2}$. Hence

$$p = \tfrac{1}{3}mn\overline{V^2} = \tfrac{1}{3}\rho\overline{V^2}$$

If v be the volume of a mass M of gas, since $\rho = M/v$,

$$pv = \tfrac{1}{3}M\overline{V^2}$$

The total kinetic energy of translation of the gas is the sum of the kinetic energies of translation of all the particles and is evidently equal to $\frac{1}{2}M\overline{V^2}$ or $\frac{3}{2}pv$. Now there is good reason to believe that, if the temperature of a gas is constant, this kinetic energy is constant. Hence the product of the pressure and volume of a gas at constant temperature is constant, and this is Boyle's Law.

In the above we have neglected collisions between particles. When two collide there is, by Newton's third law of motion, no change of the total momentum in any direction. Since the momen-

tum perpendicular to A_1 is unchanged, the pressure on it will be unchanged.

The deviations from Boyle's Law are due to the (very small) forces between particles when they are not in contact and to the fact that the molecules are not mere points but have finite sizes. These we have neglected; by considering them van der Waals arrived at his more correct law (§222).

228. Surface Condensation and Occlusion.—When a gas is in contact with a solid, there are molecular forces drawing the particles together, and these produce more or less condensation of the gas on the surface of the solid. This makes it impossible to remove the last traces of a gas from a glass vessel by means of an air pump. It also accounts for the fact that, when a figure is traced on a sheet of glass by a stick, the figure will appear when the glass is breathed on. The breath condenses less readily on the part of the glass that has been freed from condensed gas by the scraping of the stick.

A porous solid is readily permeated by a gas, and condensation on the surfaces of the pores takes place. This is called *occlusion*. Very porous wood-charcoal will absorb nine volumes of oxygen, thirty-five volumes of carbon dioxide and ninety volumes of ammonia per volume of the charcoal, and cocoanut-charcoal will absorb still more. This is why charcoal is so useful as a deodorizer and in gas masks. Platinum in sponge form will absorb 250 times its own volume of oxygen. Palladium will absorb more than one thousand volumes of hydrogen. Its own volume is thereby increased by about one-tenth. The hydrogen is therefore reduced to one thousandth of its original volume; to produce such a condensation by pressure alone would require a pressure of several tons per square inch.

229. Diffusion of Gases.—Gases, because of their greater mobility, diffuse much more rapidly than liquids. When two vessels containing different gases are connected by a wide tube, diffusion proceeds with great rapidity, and in a short time each gas is found distributed in both vessels, as if the other gas were not present. If one of the gases be a colored gas, such as chlorine, the process of diffusion can be observed. As regards the final result, each gas acts to the other as a vacuum, but in the process of diffusion each gas retards the other. Gravity also plays some part in the process, though not in the final result. Thus, if the gases be hydrogen and carbon dioxide, the final mixture is attained more rapidly when the carbon dioxide is in the higher vessel.

In the process of diffusion of two gases into each other, each gas diffuses from places where the concentration of that gas is great to places where it is less, and the rate of diffusion is proportional to the gradient of concentration (§215). It also varies

inversely as the square root of the density at a given pressure (as might be inferred from the formula $p = \frac{1}{3}\rho\overline{V^2}$ in §227).

230. Efflux of Gases.—The rate of escape of a gas through a small aperture in a very thin plate may be deduced from the principle of energy. Each part of the gas, as it escapes, has a certain velocity and therefore a certain kinetic energy, and this must equal the work performed by the pressure in the vessel in forcing the gas out. Let P be the excess of the pressure in the vessel over the external pressure. During the escape of a small volume V of the gas the pressure P does the same amount of work as if it had pushed out a piston in a cylinder. Hence (§195) the work done is PV. If the density is ρ, the mass of the volume V of the gas is $V\rho$, and if its velocity is v, its kinetic energy is $\frac{1}{2}V\rho v^2$. Equating the work done to the kinetic energy which it produces, we get

$$v = \sqrt{\frac{2P}{\rho}}.$$

Thus the rate of escape is directly as the square root of the pressure and inversely as the square root of the density.

Bunsen's method of comparing the densities of gases consists in comparing their rates of escape through the same aperture under the same pressure.

231. Diffusion through Pores.—When a gas escapes through a porous partition in which the pores are very small, such as fine unglazed pottery-ware, the circumstances are different from those of the above cases. The pores are comparable in size with the molecules of the gas, and the velocities of the molecules of different gases, at the same temperature, are different. As might be expected, the rates of escape of different gases are so different that the constituents of a mixed gas escape at different rates. This affords a method of partially separating the constituents of a mixed gas, and, as the process may be repeated several times, the separation may be made nearly complete. By this method it has been found possible to effect a partial separation of the isotopes of a gas (§159).

232. Pumps for Liquids.—The oldest form of pump, the *suction-pump*, consists of a piston moving in a cylinder or barrel that is connected with the well by a pipe. In the pipe, or at the top of the pipe, there is a valve, called the *inlet valve*, which can open towards the cylinder, but not in the opposite direction; and in the piston there is a valve, called the *outlet valve*, which can open outward but not inward toward the cylinder. When the piston

is first raised, the air in the cylinder expands, and its pressure diminishes. The outlet valve closes, owing to the excess of pressure on the outside, and, for the same reason, the inlet valve opens, and air from the suction pipe enters the cylinder Thus the air in the suction pipe is rarefied, and the greater atmospheric pressure on the water in the well forces water some distance up the suction-pipe. After some strokes the water enters the cylinder and flows out by the outlet valve.

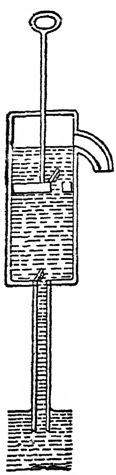


FIG. 130—Suction-pump.

Since it is the pressure of the atmosphere that raises the water in a suction-pump, water cannot be raised by this means higher in the pipe than atmospheric pressure will raise water in a vacuum; and, since the density of water is to that of mercury as 1 to 13.6, it follows that the maximum theoretical height is 13.6 times the height of the mercury in a barometer or about 34 feet. The practical limit of suction-pumps is considerably less than this, owing to the presence of air in water and to the difficulty of making the contact between piston and pump air-tight When water is to be raised higher a *force-pump* is used (Fig. 131). This differs from the suction-pump in the fact that the outlet valve is not in the piston, but in a side tube connected to the cylinder near the inlet valve During each downward stroke of the piston, water is forced up this side tube, and the height that may be reached will depend on the force that can be applied to the piston and the maximum pressure that the pump will stand without breakage of some part.

The outflow from the delivery tube of a force-pump, as just described, would be intermittent; but it may be rendered more nearly continuous by means of an "air chamber," in which air, being put under pressure by the water forced in, exerts continuous pressure on the outflowing water.

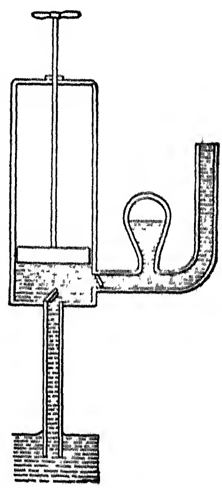


FIG. 131.—Force-pump.

233. The Siphon.—The siphon is a bent tube for removing liquid from a vessel. The tube is filled with liquid and is then inverted, and one end *A* is immersed in the liquid,

while the other end C is kept closed. When C is opened liquid flows through the tube and out through C, so long as C is below the level, D, of the surface of the liquid.

To explain the action of the siphon let us consider the pressure on the liquid at C before the end C is opened. If the difference of level of D and C is h, the pressure on the liquid at C is greater than atmospheric pressure by $g\rho h$. Hence, when C is opened, the excess of pressure inside causes a flow, and the flow continues so long as C is below the level of D and A remains immersed. A siphon will not act if the highest point B of the tube is at a greater height above the level of D than the height to which atmospheric pressure will force the liquid in an exhausted tube.

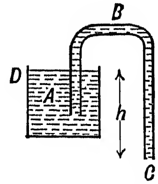


Fig. 132.—Siphon.

234. Air-pumps.—The first pump for removing air from a vessel was invented by Otto von Guericke (in 1650). It was essentially a suction pump, like that used for water, the only difference being the closer fit of piston required to prevent leakage in the case of a gas. The degree of exhaustion that can be attained by such a pump is low. The flap-valve, at the end of the suction-tube, will not act automatically, when the pressure in the receiver has become very small. For this reason a conical plug, carried by a rod that passed with some friction through the piston, was substituted. Another difficulty is caused by the fact that the piston cannot be made to fit the lower end of the cylinder with perfect accuracy, so as to expel all the air drawn from the receiver into the cylinder. The latter defect has been remedied in the Geryk pump (Fig. 133), which has a layer of oil at the bottom of the cylinder; oil above the piston also prevents leakage at the piston valve.

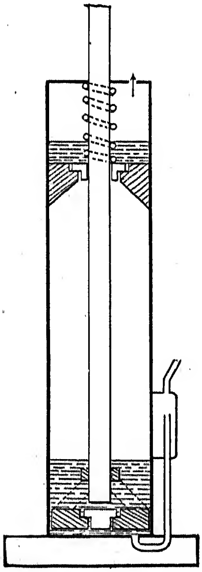


Fig. 133.—Geryk pump.

235. High-vacuum Pumps.—Many pumps were devised for removing gases from vessels and obtaining very high vacua. In nearly all of them mercury was used. In the older forms the level of mercury in a large bulb, connected to the receiver, was alternately lowered and raised, so that the gas was drawn from the

receiver into the bulb and then ejected through a side tube, or the mercury fell in drops through a narrow tube and exerted suction on a side tube connected to the receiver. These forms of mercury pump have mostly gone out of use, and we shall describe only some of the more recent and efficient types of air-pumps.

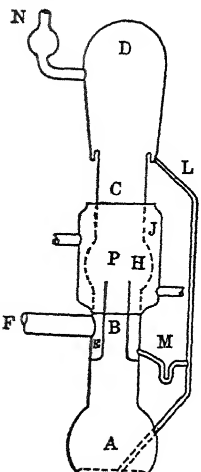


Fig 134—Langmuir's mercury vapor pump

Langmuir's "mercury vapor pump" makes use of the principle of the aspirator (§193) in a novel form. A current or "blast" of mercury vapor passes upward from the heated flask, A, through the tubes B and C (Fig. 134) into the condenser D. An annular space, E, surrounding B, is connected through F with the vessel to be exhausted. C is enlarged into a bulb, H, just above the upper end of B, and H is surrounded by a condenser, J, through which water flows. The mercury, condensing in D and H, returns to A through the tubes L and M. The gas from F passes freely up through E, and, meeting the blast of mercury vapor at P, is blown outward and upward along the walls of the condenser H and forced into the main stream of mercury vapor passing up through C into the condenser D. A less efficient pump, connected to N, maintains a vacuum of about 0.3 mm. (400 bars) and removes the gas Langmuir's pump will exhaust a vessel of 11 liters capacity from atmospheric pressure to a vacuum of 0.00001 mm. (0.015 bars) in 80 seconds. Because of its remarkable simplicity and rapidity of action, it marks a great advance in methods of obtaining high vacua.

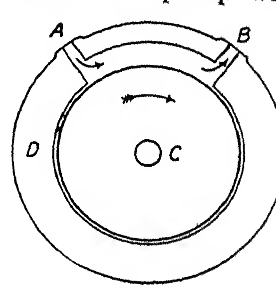


Fig. 135.

Several other forms of air-pump have been devised. Two that are particularly simple in their action may be mentioned. Fig. 135 shows the principle of Gaede's *molecular pump*. A cylinder C rotates within a cylinder casing, D. The fit is close, except between A and B. The air (or gas) is drawn in at A and dragged around by molecular impacts to B, where it is ejected. The actual structure of the pump is somewhat more complex than shown in the figure. Several such units, with

only one rotating cylinder, can be connected in series, and so a very high vacuum can be obtained. In another pump, for producing moderate vacuua (Fig. 136), a cylinder, C, rotates eccentrically within a cylindrical casing. Air drawn in at A, is chased around to B, where it is ejected. A plug, P, that moves in a slot is kept forced down on C by a spring and prevents the passage of air from B to A.

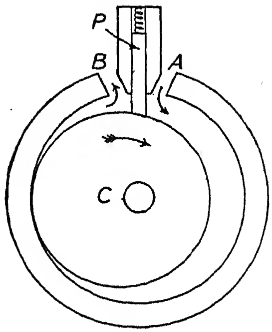


FIG. 136.

The highest vacuum obtainable by a pump can be improved by means of a tube that contains charcoal and is connected to the vessel to be exhausted. When the tube is immersed in liquid air, most of the remaining gas is absorbed by the charcoal.

PROBLEMS

Velocity and Acceleration.

1. A train acquires 5 minutes after starting a velocity of 40 km. per hour. Assuming constant acceleration, find the distance passed over in the 5th minute *Ans.* 0.6 km.
2. A train having a speed of 70 km. per hour is brought to rest by brakes in a distance of 600 m. What is the acceleration (assumed constant)? *Ans.* -1.13 km./min.^{2}
3. What is the final speed of a body which, moving with uniform acceleration travels 72 meters in 2 minutes if:
   (a) the initial speed = 0?
   (b) the initial speed = 15 cm. per sec.?
   Ans. 120 cm./sec.; 105 cm./sec.
4. The shell of a trench mortar has a muzzle velocity of 180 ft./sec. and it travels 3.5 ft. in the bore. What is its (average) acceleration in the bore? *Ans.* 4629 ft./sec.^{2}
5. A body slides down an inclined plane and in the 3d sec. describes 110 cm. What is the inclination? *Ans.* 2° 35′.
6. What initial vertical velocity must a ball have in order to fall back to its starting point in 10 sec.? *Ans.* 4900 cm./sec.
7. At what angle with the shore must a boat be directed in order to reach a point on the other shore directly opposite, if the speed of the boat be 4 miles per hour and that of the stream be 2 miles per hour? *Ans.* 60°.
8. A point goes over a circular path 10 cm. in diameter 4 times a second, at a uniform speed. To what acceleration is it subject? *Ans.* 3158 cm./sec.^{2}
9. A baseball rises to a height of 50 ft. and travels 200 ft. horizontally before it reaches the ground. Find the direction and magnitude of the velocity with which it is thrown. *Ans.* $\theta = 45°$; $v = 80.2$ ft/sec.

10. A body is projected at an angle of 30° with the horizon with a velocity of 30 m. per sec. When and where will it again meet this horizontal plane? How far will it ascend? *Ans.* 3.06 *s.*; 79 5 m ; 11.4 m.

11. A trench mortar throws a shell with a muzzle velocity of 180 ft./sec. When the elevation is 40° the range is 295 yards. How much is the range diminished by air resistance? *Ans.* 35 yards.

12. What force will a man who weighs 70 kg. exert upon the floor of an elevator descending with an acceleration of 100 cm per sec. per sec ? If ascending with the same acceleration? *Ans.* 62.8 kg. wt.; 77.1 kg. wt.

Force and Mass.

13. A force of 1000 dynes acts upon a mass of 1 kg. for 1 min. Find the velocity acquired and the space passed over in this time. *Ans* 60 cm./sec.; 1800 cm.

14. A shot weighing 10 lbs. is shot from a gun weighing 3 tons with an initial velocity of 1200 feet per sec. What is the initial velocity of the recoil? *Ans* 2 ft./sec.

15. Three forces, 5, 12, 15 are in equilibrium. Find the angles between them. *Ans* 62° 11′; 134° 58′; 162° 51′.

16. Bodies of mass 10 kg. and 8 kg. are connected by a string over a pulley. How far does each move from rest in the first two seconds? *Ans.* 218 cm.

17. The shell of a trench mortar weighs 12 lbs It travels 3 5 ft. in the bore and has a muzzle velocity of 180 ft./sec. What is the average force acting on the shell while it is traveling in the bore? *Ans.* 1725 lbs. wt

18. An observer in an airplane noticed that at the top of a slow loop objects just began to drop toward the earth. Assuming the radius of curvature to be 40 ft. what was the speed of the airplane? *Ans.* 24 5 mi./hr

19. A baseball whose mass is 300 g. when moving with a velocity of 20 m per sec is squarely struck by a bat and then has a velocity of 30 m. per sec. in the opposite direction. Calculate the impulse and average force, if the contact lasts .02 sec. *Ans* 15×10^5 g. cm./sec.; 7.5×10^7 dynes.

20. With how much energy must a bullet weighing 20 g. be shot horizontally from a gun 4 m. above a level plane, in order to strike the ground 300 m. away from the gun? *Ans.* 1.11×10^{10} ergs.

Work and Energy.

21. A projectile traveling at the rate of 700 ft. per sec. penetrates to the depth of 2 in. Find the velocity necessary to penetrate 3 in. *Ans.* 857 ft./sec.

22. A hammer of 6 kg. mass and moving with a velocity of 100 cm. per sec. drives a nail into a plank 1 cm. What resistance does it overcome (supposed uniform)? *Ans.* 3×10^7 dynes.

23. A man can bicycle 12 miles an hour on a smooth road; downward force of each foot in turn = 20 lbs., length of stroke = 1 ft., bicycle is advanced 17 ft. for each revolution of the cranks. At what H.P. does he work? *Ans.* .075 H.P.

24. A man who weighs 180 lbs. runs up 26 steps, each 7 in. high, in 4 sec. At what H.P. does he work? *Ans.* 1.2 H.P.

25. A sprinter who weighs 161 lbs. runs 40 yds. in $4\frac{4}{5}$ sec., 60 yds. in $6\frac{4}{5}$ sec., 100 yds. in 10 sec. What is (a) his velocity from 40 to 60 yds and from 60 to 100 yds., (b) his kinetic energy at the end of 40 yards? (c) Calculate

the rate of working in H.P. required to produce this kinetic energy. (d) In what other ways does he expend energy?

Ans. (a) $33\frac{1}{3}$ ft./sec. (b) 2777 ft. lbs. (c) 1.09 H.P.

26. Find the number of watts in one horse-power. *Ans.* 746.

27. A sprinter does 100 yards on the horizontal in 10.5 sec., and the same distance up hill with a rise of 32 ft. in 17.5 sec. Assuming that his rate of working is the same throughout, calculate the added work done in the additional 7.0 seconds up hill and the rate of working that this implies, the man's weight being 160 lbs. *Ans.* 1.33 H.P.

28. 100 cu. ft. of water pass over a dam 10 ft. high in 1 min. What horse-power could be derived from this if all were utilized? *Ans.* 1.9 H.P.

29. A 30-gram rifle bullet is fired into a suspended block of wood weighing 15 kg. If the block is suspended by a string of length 2 meters and is moved through an angle of 20°, calculate the velocity of the bullet. Notice that the impact of the bullet on the block does not change the total *momentum* of both (§47) and during the subsequent swing of the pendulum its total *energy* remains constant. *Ans.* 770 m./sec.

30. If a locomotive driving-wheel 1.5 m. in diameter makes 250 revolutions per minute, what is the mean linear velocity of a point on the periphery? Of the point when it is highest? When it is lowest?

Rotation.

Ans. 19.6 m./sec.; 39.2 m./sec.; 0.

31. The armature of a motor revolving at the rate of 1800 revolutions per minute comes to rest in 20 seconds after the current is shut off. Calculate its average angular acceleration and the number of revolutions in this time.

Ans. -9.42 rad./sec.^{2}; 300 rev.

32. Find in radians per second the angular velocity of the earth about its axis and deduce the component of this angular velocity about a diameter through a point in latitude 40°. (Principle of Foucault's pendulum).

33. A circle has a diameter of 16 cm. A smaller circle tangent to it and 12 cm. in diameter is cut out of it. Where is the center of gravity of the remainder?

Ans. 10.6 cm. from common tangent.

Center of Mass.

34. Two cylinders of equal length (=20 in.), and having diameters of 12 and 6 in., are joined so that their axes coincide. Where is the center of gravity? *Ans.* 6 in. from junction.

35. Find the center of gravity of a table 4 ft. × 3 ft. × 1 in., with legs at the corners 2 ft. × 2 in. × 2 in. *Ans.* 0.233 ft. from top.

36. The mass of the moon is $\frac{1}{80}$ of that of the earth and the average distance between their centers is 240,000 miles. Calculate the position of the center of mass of the two. *Ans.* 2963 m. from center of earth.

37. At the corners of an equilateral triangle ABC masses of 1, 2 and 3 lbs. respectively are placed. Find the distance of their center of mass from BC, assuming each side of the triangle to be 1 ft. in length. *Ans.* 0.144 ft.

38. A bar 6 ft. long and pivoted at the middle has a weight of 24 lbs. hung at one extremity. What is the moment of the weight (a) when the bar is horizontal, (b) when it makes an angle of 30° below, and (c) of 60° above with the horizontal?

Moments.

Ans. 72; 62.3; 36 lbs. wt. ft.

39. If it is wished to upset a tall column by a rope of given length pulled from

the ground, where should it be applied, if the length of the rope is, (1) equal to, (2) twice, the height of the column?

40. Find the rotational inertia of a sphere ($m = 20$, $r = 2$) about an axis tangent to its surface. *Ans.* 112.

41. Find the rotational inertia of three circular disks, all three touching each other in the same plane, about a perpendicular axis passing through the center of one of them. The mass of each is 100 g. and the radius of each is 6 cm. *Ans* 34,200 gm cm.^{2}

42. Two masses, 1 kg and 2 kg., respectively, are connected by a rod of negligible mass, 1 m long. The system is thrown so that the center of gravity has a velocity of 20 m per second and the system turns 10 times per second about this center Find the kinetic energy of the system. *Ans.* 192×10^8 ergs.

43. What energy has a grindstone $1\frac{1}{2}$ m. in diameter, weighing 1000 kg and rotating once every 2 sec ? *Ans* $13\,9 \times 10^9$ ergs.

44. A solid iron cylinder, 100 cm diameter, rolls down a plane 6 m long inclined at 30°. What linear velocity does it acquire? *Ans* 627 cm./sec.

45. A block of stone weighs 2 5 tons and is in the form of a cube of 1 yard side. It rests on level ground What is the least force which applied to the block will cause it to revolve about a horizontal edge? *Ans.* 1768 lbs. wt.

46. Parallel forces of 1, 2 and 3 units respectively act at the corners A, B, C of an equilateral triangle of 1 ft. side. Find the distance of the resultant from BC *Ans.* 0.144 ft.

Parallel forces. 47. Parallel forces of 10 and 6, but in opposite directions, are applied to a bar at distances of 8 and 3 from one end. What is the magnitude of the resultant and where does it act? *Ans.* 4; 15.5.

48. Two equal parallel forces, each 50 dynes, act in opposite directions at the ends of a bar 10 cm. long. The bar makes an angle of 45° with the direction of the forces. What is the moment of the couple? *Ans.* 353 5 dynes-cm.

49. A man and a boy carry a weight of 20 kg between them by means of a uniform pole 2 m. long, weighing 5 kg Where must the weight be placed so that the man may carry twice as much of the whole weight as the boy? *Ans.* 0 416 m. from middle.

50. A rod, the mass of which is 1 kg., hangs from a hinge on a vertical wall and rests on a smooth floor. Calculate the force on the floor and the force on the hinge. *Ans.* 500 g.; 500 g.

Equilibrium. 51. A uniform ladder 30 feet long and of 80 lbs. weight rests with the upper end against a smooth vertical wall, and the lower end is prevented from slipping by a peg If the inclination of the ladder to the horizontal is 30°, find the force on the wall and at the peg. *Ans.* 43.3 lbs. wt.; 66.1 lbs wt.

52. A barn door is 10 ft. long and 5 ft. wide and weighs 200 lbs. The hinges are 1 ft. from the ends and the weight is carried entirely by the upper hinge. Find the direction and magnitude of the resultant force on the upper hinge. *Ans.* 209 lbs wt ; 17° 21′ to vertical.

53. One end of a certain rod is clamped. If the other end is pulled 1 cm. from its natural position and then released, it starts with an acceleration of 10 cm. per sec. per sec. What is the period of its vibration?

Periodic Motions.

Ans. 1.98 sec.

54. The balance-wheel of a watch makes 5 complete vibrations in 2 sec. With what angular acceleration will it start when turned 30° from its position of equilibrium and released? *Ans.* 129.34 rad./sec.^{2}

55. A hoop of 25 cm. radius hangs on a peg. Prove that its period of vibration is equal to that of a simple pendulum whose length is equal to the diameter of the hoop.

56. A clock gains 3 min. a day. Find the error in the length of the pendulum, regarding it as a simple pendulum. (g = 980.) *Ans.* 0.414 cm.

57. A pendulum which is a "second's pendulum" where g = 980, vibrates but 59.95 times a minute on top of a mountain. What is the acceleration of gravity at this point? *Ans.* 978.37.

58. A rod 2 m. long is freely suspended at one end. Calculate its period of vibration. *Ans.* 2.31 sec.

59. A "second's pendulum" is drawn aside and released and at the same moment a ball is allowed to fall. The ball and the bob collide as the pendulum passes through the vertical. Calculate the height of fall of the ball. *Ans.* 122.5 cm.

60. The coefficient of friction for two surfaces = 0.14. A pull of 20 kg. weight will overcome what pressure between them? *Ans.* 143 kg.

Friction.

61. What force applied parallel to a plane inclined at 20° will push up a block weighing 100 kg., the coefficient of friction between the two being 0.24: (*a*) the block moving uniformly; (*b*) the block having an acceleration of 100 cm. per sec. per sec.?

Ans. (*a*) 56.7 kg. wt.; (*b*) 66.9 kg. wt.

62. What is the coefficient of friction between a body and a horizontal plane if the body loses a velocity of 100 ft. per sec. and comes to rest in moving 200 ft. over the plane? *Ans.* 0.776.

63. A toboggan slides 100 yards down a track inclined at 20° to the horizontal in 11 seconds. Calculate the coefficient of friction. *Ans.* 0.20.

64. A small block rests on a horizontal revolving platform at a distance of 40 cm. from the axis of revolution. If the coefficient of friction is .30 at what angular velocity of the platform will the block just begin to slip? *Ans.* 2.71 rad./sec.

65. A man raises a stone 1 in. with a lever of the first class 10 ft. long weighing 50 lbs., the fulcrum being 1 ft. from the point of application to the stone. If he exerts a force of 100 lbs. wt. what force is applied to the stone and what work does he do?

Machines.

Ans. 1,100 lbs. wt.; 75 ft. lbs.

66. A boy who exerts a push of 50 lbs. wt. wishes to roll a barrel weighing 200 lbs. into a wagon 2½ ft. high. Assuming that he pushes in a line through the center of the barrel parallel to the plank, how long a plank will he need and how much work will he do? *Ans.* 10 ft.; 500 ft. lbs.

67. A body weighs 12 lbs. on one side of a false balance and 12.5 lbs. on the other side. What is the ratio of the arms of the balance? *Ans.* 1.021.

68. A man weighing 150 lbs. sits on a platform suspended from a movable pulley and raises himself by a rope passing over a fixed pulley. Supposing the cords are parallel, what force does he exert? *Ans.* 50 lbs. wt.

69. A wheel whose radius is 25 cm. is fastened to one end of a screw whose pitch is 1 mm. What force can the screw exert in its nut when a force of 1 kg. wt. is applied tangentially to the wheel, friction being supposed negligible? *Ans.* 1570 kg. wt.

70. Compare the mechanical advantages of a block and tackle when the end of the cord is attached to the upper block and when it is attached to the lower.

71. How far above the surface of the earth must a body be to lose 0.10 per cent. in weight? *Ans.* 2.0 mi.

Gravitation. 72. If the moon's mass is $\frac{1}{80}$ that of the earth, and its diameter 2160 miles, that of the earth being 7900 miles, what is the acceleration of gravity on the moon's surface? *Ans.* 164 cm./sec.^{2}

73. Find the time of revolution of the earth which would cause bodies to have no apparent weight at the equator *Ans.* 1.41 hr.

74. A wire 300 cm long and 1 mm. in diameter is stretched 1 mm. by a weight of 3000 g What is Young's Modulus? *Ans.* 11.2×10^{11} dynes/cm.^{2}

Elasticity. 75. A weight is hung from the ceiling by a steel wire 2 m. long and of 1 mm. diameter joined to a copper wire 1 m. long and of 0.5 mm in diameter. Another weight sufficient to produce a total extension of 1 mm is added. Calculate the extension of each part. *Ans.* 0.19 mm.; 0.81 mm.

76. To opposite faces of a cubical block of jelly of 20 cm edge parallel and opposite forces of 1 kg. each are applied and produce a relative motion of 1 cm Calculate the strain, the stress and the shear modulus. *Ans.* 0 05; 2450 dynes/cm.^{2}; 49,000 dynes/cm.^{2}

77. An iron bar of 400 c.c. volume falls from a ship and sinks to the bottom of an ocean 1000 m. deep. How much is its volume diminished, assuming that each 10 m. of water pressure produces a pressure equal to that of the atmosphere, which equals one million dynes per sq. cm.? *Ans.* 0.026 c.c.

78. A ball weighing 20 kg., moving with a velocity of 500 cm. per sec, strikes a second ball weighing 100 kg. which is at rest. If the first ball rebounds with a velocity of 100 cm per sec., what will be the velocity of the second? *Ans.* 120 cm./sec.

79. Two bodies differing in bulk weigh the same in water, compare their weights in mercury; in vacuo.

Properties of Liquids. 80. A mass of copper suspected of being hollow weighs 523 g. in air and 447 5 g in water. What is the volume of the cavity? *Ans.* 16.8 c.c.

81. The specific gravity of ice is 0.918, that of sea-water 1.03. What is the total volume of an iceberg of which 700 cu. yds. are exposed? *Ans.* 6438 cu. yds.

82. A block of wood weighing 1 kg., whose specific gravity is 0.7, is to be loaded with lead so as to float with 0.9 of its volume immersed. What weight of lead is required, (1) if the lead is on top, (2) if the lead is below? *Ans.* 286 g; 313.5 g.

3. A hydrometer sinks to a certain mark in a liquid of sp. gr. 0.6, but it takes 120 g. to sink it to the same mark in water. What is the weight of the hydrometer? *Ans.* 180 g.

4. One of the limbs of a U-shaped glass tube contains mercury to the height of 0.175 m., the other contains a different liquid to a height of 0.42 m., the two columns being in equilibrium. What is the specific gravity of the second liquid with reference to mercury and to water.

5. Find the volume in cu. ft. of the smallest block of ice which, floating on fresh water, will just carry a man who weighs 150 lbs. *Ans.* 29.3 cu. ft.

6. Given a body A which weighs 7.55 g. in air, 5.17 g. in water, and 6.35 g. in another liquid *B*, find the specific gravity of the body *A* and that of the liquid *B*. *Ans.* 3.17; 0.504.

7. A block of brass 10 cm. thick floats on mercury. How much of its volume is above the surface, and how many cm. of water must be poured above the mercury so as to reach the top of the block? (Density of mercury = 13.6; of brass = 8.5.) *Ans.* 0.375 of the whole; 4.05 cm.

8. Two tubes are inserted in a vessel of water on the same horizontal plane. The diameter of the one is 0.5 mm. and its length is 20 cm.; the diameter of the other is 0.25 mm. and its length is 10 cm. Compare the amounts of water flowing through the two tubes in a given time. *Ans.* 8:1.

9. The diameter of the small piston of an hydrostatic press is 2 in., the diameter of the large piston is 2 ft. What weight on the small piston will support two tons on the large piston? *Ans.* 27.77 lbs.

0. The pressure at the bottom of a lake is three times that at a depth of 2 m. What is the depth of the lake? (Atmospheric pressure = 76 cm. of mercury.) *Ans.* 26.67 m.

1. A retaining wall 3 m. wide and 40 m. long is inclined at 30° to the horizontal. Find the total force in kg. exerted against it by the water when the water rises to the top. *Ans.* 9×10^4 kg.

2. What is the outward force exerted by the water on the sides of a circular tank 1 m. in diameter, the height of the water being 150 cm.? What is the thrust due to the water on the bottom? *Ans.* 3532 kg. wt.; 1178 kg. wt.

3. The surface tension of a soap-bubble solution is 27.45 (dynes/cm.). How much greater is the pressure inside a soap-bubble of .3 cm. radius than in the air outside? *Ans.* 36.6 dynes/cm.^{2}

4. How far will water be projected horizontally from an aperture 3 m. below the water level of a tank and 10 m. above the ground (neglecting air resistance)? *Ans.* 10.96 m.

5. A body whose specific gravity is 2 is weighed in air of specific gravity 0.0013 with weights of specific gravity 9. The weight in air being 100 g., what is the true weight? *Ans.* 100.050 g.

Properties of Gases

96. If the barometer sinks 15 mm., how much is the pressure in dynes per sq. cm. decreased? *Ans.* 19992 dynes/cm.^{2}

7. An air bubble at the bottom of a pond 6 m. deep has a volume of 1 c.c. Find the volume just as it reaches the surface, the barometer standing 760 mm. *Ans.* 1.58 c.c.

98. Owing to the presence of air the mercury column in a barometer 85 cm. long stands at 70 cm. when an accurate barometer stands at 75 cm. What pressure will this barometer indicate when an accurate barometer stands at 72 cm.? *Ans.* 67.67 cm.

99. A barometer reads 73 cm. Calculate the thrust on one side of a board 1 m. square. *Ans.* 9928 kg. wt.

100. A barometer has a cross-section of 2 sq. cm. and is so long that, as the mercury stands at 76 cm., there is a vacuum space 10 cm. long. Some air is allowed to enter and the mercury falls 10 cm. What was the volume of the air before it entered? *Ans.* 5.26 $cm.^3$

101. How high must we ascend above the sea-level to observe a depression of 1 mm. in the height of the barometer? Density of air = 0.0013 (approx.). *Ans.* 10.4 m.

102. A glass tube 60 cm. long, closed at one end, is sunk, open end down, to the bottom of the ocean. When drawn up it is found that the water has penetrated to within 5 cm. of the top. Atmospheric pressure = 76 cm. of mercury. Calculate the depth of the ocean, assuming the density constant, and equal to 1.026. (Principle of Lord Kelvin's sounding apparatus.) *Ans.* 110.8 m.

103. In a vessel of 1 cu. meter volume are placed the following amounts of gas: (1) hydrogen, which occupies 1 cu. m. at atmospheric pressure. (2) nitrogen, which occupies 3 cu. m. at a pressure of 2 atmospheres. (3) oxygen, which occupies 2 cu. m. at a pressure of 3 atmospheres. Calculate pressure of mixture. *Ans.* 13 at.

104. The mouth of a vertical cylinder 18 in. high is closed by a piston whose area is 6 sq. in. If a weight of 100 lbs. be placed on the piston, how far will it descend, supposing the atmospheric pressure to be 14 lbs. per sq. in., the friction negligible and the temperature constant? *Ans.* 9.8 in.

105. A cylindrical diving-bell 7 ft. in height is lowered until the top of the bell is 20 ft. below the surface of the fresh water. If the barometric height at the time is 30 in., how high will the water rise in the bell? What air pressure in the bell would just keep the water out? *Ans.* 2.96 ft.; 1.82 at.

106. (a) What fraction of an atmosphere is the difference in pressure between two points in air at 0° C. and 76 cm. pressure if the difference of level is 1 cm.? (b) How large a difference of level would produce a difference of pressure of 0.01 per cent. of an atmosphere? *Ans.* 126×10^{-6}; 80 cm.

DIMENSIONS OF MECHANICAL UNITS

Linear velocity, v	LT^{-1}	Moment of Force, L	$L^2T^{-2}M$
Linear acceleration, a	LT^{-2}	Rotational Inertia, I	L^2M
Angular velocity, ω	T^{-1}	Work, W	$L^2T^{-2}M$
Angular acceleration, α	T^{-2}	Kinetic Energy, E	$L^2T^{-2}M$
Force, F	$LT^{-2}M$	Potential Energy	$L^2T^{-2}M$

WAVE MOTION

By E. Percival Lewis, Ph. D.

Late Professor of Physics in the University of California

Revised[1] by F. A. Jenkins, Ph. D.

University of California

236. Characteristics of Wave Motion.—The word **wave** recalls the familiar phenomena observed whenever the surface of a body of water is disturbed. Large waves are usually so irregular that it would be difficult to reach any general conclusions regarding the laws of their formation or propagation. If less complex waves be observed, such as those produced by throwing a pebble into a quiet pond or by the gentle disturbance of the water or mercury in a tank, it will be seen that they consist of alternate ridges and hollows in the surface, which diverge in uniformly expanding circles from the center of disturbance. If small pieces of cork rest on the surface another important characteristic of wave motion may be observed. The particles rise on an approaching wave, ride forward on its crest for a short distance, then fall into the succeeding hollow with a backward motion, to again move upward and forward on the next crest. They describe orbits in a vertical plane which are evidently circular or elliptical. Since these particles participate in the movement of the water on which they rest, it is plain that the water as a whole does not move continuously forward with the waves, but that each element rotates about its original undisturbed position, to which it returns when the train of waves has passed. *Waves are, therefore, the progression of a shape or condition.*

Water waves illustrate the following fundamental characteristics of all wave motions in material media:

(1) *All parts of the medium reached by the disturbance are subject to periodic displacements about their positions of equilibrium.*

[1] The revision of Wave Motion and Light for the seventh edition was by R. T. Birge and E. E. Hall, University of California, and valuable features of their work are incorporated in the present revision.

(2) *The disturbance is propagated at a uniform rate, each displaced particle transferring its motion to its neighbors by pressure or through some mechanical connection.* The moving elements of the medium possess kinetic energy due to their motion and potential energy due to their displacements. This energy, originally derived from the source of disturbance, is passed on from element to element so that there is a continuous *flow of energy* with the advancing waves.

237. Transverse Waves.—The displacements in the case of water waves do not extend far beneath the surface, hence disturbances are propagated in two dimensions only, in the form of superficial waves. There is another familiar type, resembling water waves in general shape, which may be propagated along a linear medium, such as a wire or rope. These may be called linear waves (although the disturbance extends across a finite area) because they are propagated in one direction only. Such waves may be studied by filling a long rubber tube with shot and suspending it from a tall support, holding the lower end taut in the hand. If the tube is struck a sharp blow near the lower end, a distortion resembling a wave crest will travel slowly to the upper end, where it will immediately be reflected with reversed curvature, on account of the elastic reaction of the fixed point. (Fig. 137, *a*, *b*, *c*.) It will then travel to the lower end and be reflected back and forth several times until its energy is exhausted by friction. This is a *solitary* wave. If the lower end is rapidly moved back and forth through a small amplitude, by properly timing the displacements it is possible to cause a series or *train* of waves, consisting of "crests and hollows," to travel upward, crossing a similar train reflected downward. The combined effect of the two trains is to cause the tube to oscillate between the positions shown by the full and the dotted line in Fig. 137, *d*.

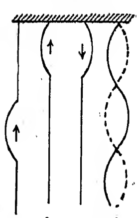

a b c d

FIG. 137.

In the cases of the waves so far mentioned the oscillations of the medium are in part or wholly at right angles or *transverse* to the direction of propagation, so that the displacements of the boundary of the medium give rise to a definite wave shape.

238. Longitudinal Waves.—It is possible for the vibrations in a wave to take place in the direction of propagation of the wave, as is the case with one component of the displacement in water waves. When the displacements are altogether in the direction of propagation it is evident that the wave can have no shape,

as the boundaries of the medium are not displaced, but there will be periodic changes in density, arising from the fact that different particles are at any instant in different *phases* of displacement, so that in one region they will be crowded together, while in another they will be separated. This may be illustrated by a row of massive spheres, connected by elastic cords or springs, as shown in Fig. 138, *a*. If the second sphere were immovable, the first alone would oscillate when pulled downward and released. If the spheres are all free to move the transmitted impulse will set all in vibration. On account of the inertia of the spheres and the elasticity of the connections, the displacement of each sphere will lag behind that of its neighbor below, and each vibration will be in a different phase, until we come to the sphere *B* (Fig. 138 *b*), which begins its first vibration when *A* begins its second vibration. The figure shows the resultant effect when the first sphere has completed one vibration (*b*) and one and a half vibrations (*c*) after it first moved upward through its resting point. It is evident from the figure that the *conditions* of condensation and of rarefaction are propagated with the velocity of the wave. There is no change of shape in the system, but if lines proportional to the displacements are drawn from each resting point, to the right for upward displacements, to the left for downward displacements (that is, if each displacement is rotated through 90° to the right or the left), a smooth curve drawn through the ends of these lines will have the general shape of a transverse wave (*b*, *c*). We have thus a means of graphically representing *longitudinal* waves in a way clearly coordinating them with *transverse* waves.

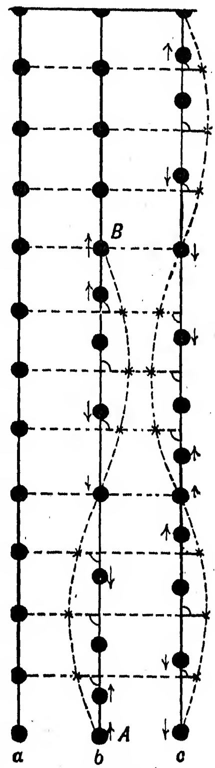


Fig. 138.

239. Other Types of Waves.—If a series of heavy bars are attached horizontally at equal intervals to a suspended wire, and if the lowest bar executes torsional vibrations, waves of angular displacement will travel up the wire. Such *torsional* waves may be represented graphically by erecting ordinates proportional to the angle of torsion at each point on an axis representing the wire.

There are many cases where wave disturbances, such as those of sound in air, are propagated in three dimensions in a uniform medium. These disturbances will travel equal distances in all directions in equal times, hence the waves will be *spherical*, with the source as a center. A hemispherical wave of this type would be produced in a large block of metal by striking it at a point.

So far we have considered only the effect of mechanical disturbances of a medium. The idea of wave motion may, however, be extended to cases where any physical condition in a medium varies periodically at each point and is propagated with a finite velocity through the medium. A familiar example is found in the "heat waves" which travel into the earth as a result of the periodic heating and cooling of the surface. In the afternoon the surface reaches a maximum temperature. Owing to the slow conduction of the heat, this maximum travels slowly downward, the amplitude continually decreasing, owing to the fact that each particle passes on only a portion of the energy received by it, not nearly all as in the case of elastic media. At night the surface reaches a minimum temperature which penetrates into the soil at the same rate as the maximum. The distribution of temperatures in the afternoon and at night are represented by the full and the dotted line in Fig. 139. The abscissa of the point A represents the average temperature. AB is the distance traveled by the heat wave in twenty-four hours.

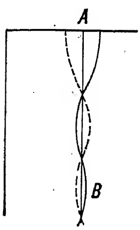


FIG. 139.

Another example of non-material waves is found in the electrical waves traveling along conductors or in free space, due to periodic change in the electrical condition at different points. Light also consists of electrical waves of very short wave-length (§523).

240. Vibrations in Wave Motion.—In all departments of physics, waves play an important part; hence the study of wave motion is of fundamental importance. Since periodic displacements or changes in condition are an essential feature of wave motion, it is necessary to study such phenomena in detail. Now the only periodic motions which lend themselves readily to simple analysis are those of uniform motion in a circle and its projections along a line, the latter being called simple harmonic motions (§114 *et seq.*). As pointed out in §117, the vibrations of all elastic bodies must be either simple harmonic motions or compounded of such

motions (§248), since, for small displacements at least, the restoring force is proportional to the displacement.

Since a simple harmonic motion is a changing linear displacement, it may be resolved into two or more components like any displacement (§14). If, for example, the piston rod AB (Fig. 140) executes simple harmonic vibrations in a horizontal line (the projected motion of the crank pin on a fly-wheel), a pin P attached to it and sliding in a slotted cross bar attached to the rod CD will cause the latter to execute a simple harmonic vibration in the direction of its length, if guides allow it to move only in that direction. If the amplitude of AB is r, the length of the crank arm, that of CD is $r \cos \alpha$. This gives one component of the simple harmonic motion of AB, and the other (not shown by the apparatus) is at right angles to CD.

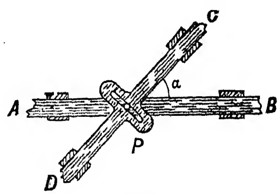


Fig. 140.

241. Superposition of Simple Harmonic Motions.—In many cases a body may be subjected to several simultaneous simple harmonic displacements in the same or in different directions and of the same or different periods. Familiar illustrations are found in the vibrations of musical instruments (§581 *et seq.*) and whenever different sets of waves are superimposed on or cross each other. If the displacements are entirely independent, it is evident that the resultant effect may be obtained by the geometrical addition of displacements (§13). If a light pendulum is suspended from a heavy one, as shown in Fig. 141, and both set in vibration in the same plane, at a given instant the total displacement of the lower bob is $x = x_1 + x_2$; or if the pendulums vibrate at right angles, the resultant displacement is $r = \sqrt{x^2 + y^2}$. In such a case the two systems are not entirely independent, on account of their connections and inertia, and the two displacements will not remain of the simple harmonic type. If a simple pendulum be set in vibration, and later an impulse at right angles to its direction of motion be applied, it will move in a circular or elliptic orbit (conical pendulum), or in a line inclined to its original direction. In studying these effects the most useful cases to consider are those in which the periods of the components are either equal or in some simple ratio to one another.

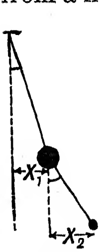


Fig. 141.

242. Composition of Two Simple Harmonic Motions of Same Period and in Same Line.—Let us assume that a body at O (Fig. 142) is simultaneously subjected to two vibrations, each a simple harmonic motion, given by the equations

$$x_1 = r_1 \cos (\omega t + e_1) \tag{1}$$
$$x_2 = r_2 \cos (\omega t + e_2) \tag{2}$$

The resulting displacement, X, of the body at time t is the algebraic sum of x_1 and x_2. An expression for X can be found from Fig 142. For x_1 can be considered as the projection on the x-axis (horizontal) of $OC_1(=r_1)$, which revolves about O with the uniform angular

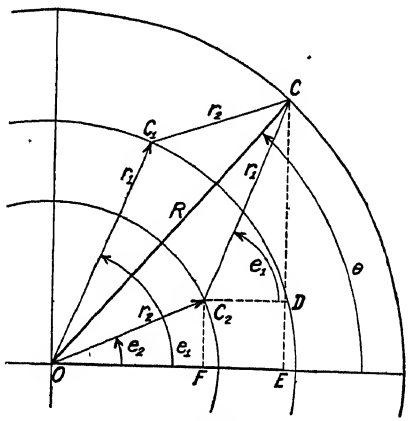


Fig 142.

velocity ω, and makes an angle e_1 with the x-axis at time $t = 0$. Similarly x_2 is the projection of $OC_2(=r_2)$ which revolves at the *same* rate ω, and initially makes an angle e_2 with the x-axis. Hence the *relative* phase of x_1 and x_2, which is the same as that of C_1 and C_2 (§120), equals $(e_1 - e_2)$ and is *constant.*

Complete the parallelogram OC_2CC_1. This entire parallelogram revolves about O with the angular velocity ω. Its diagonal R makes an angle θ with the x-axis, at time $t = 0$, and an angle $(\omega t + \theta)$ at any other time t. The projection, X, of R is, by geometry, equal to the sum of the projections of OC_1 and OC_2, or $(x_1 + x_2)$, and

$$X = R \cos (\omega t + \theta) \tag{3}$$

Hence the resultant motion, as given by (3) is **a simple harmonic motion of the same period as that of the components,** *of amplitude* R *and with an initial phase* θ *intermediate between the phases of the components.* It is the projected motion of the point C in the resultant circle of reference of radius R.

To determine X we must obtain R and θ as functions of the given quantities r_1, r_2, e_1 and e_2. From the figure

$$R^2 = r_1^2 + r_2^2 + 2r_1r_2 \cos (e_1 - e_2) \tag{4}$$

$$\tan \theta = \frac{CE}{EO} = \frac{CD + DE}{EF + FO} = \frac{r_1 \sin e_1 + r_2 \sin e_2}{r_1 \cos e_1 + r_2 \cos e_2} \tag{5}$$

It is evident from (4) that R is a maximum, $(r_1 + r_2)$, when $(e_1 - e_2) = 0$, and R is a minimum, $(r_1 - r_2)$, when $(e_1 - e_2) = 180°$.

While the above refers to the addition of two simple harmonic motions of the same period, we can extend it to the case of two vibrations of different periods by supposing the phase difference, $e_1 - e_2$, to change uniformly with the time. We may suppose the two motions to start at the same instant, $e_1 - e_2$ being then 0. At time t, the value of $e_1 - e_2$ will be $(\omega_1 - \omega_2)t = 2\pi(n_1 - n_2)t$, where n_1 and n_2 are the respective frequencies. When $(n_1 - n_2)t =$ 0, 1, 2, 3, etc., $\cos (e_1 - e_2)$ will be 1 and, from (4), R will be a maximum $(r_1 + r_2)$. When $(n_1 - n_2)t = \frac{1}{2}, \frac{3}{2}$, etc., R will be a minimum $(r_1 - r_2)$. The interval between two successive maximum values of R is $1/(n_1 - n_2)$ and the number of maxima per second is $(n_1 - n_2)$. This case is illustrated by "beats" in Sound (§578).

243. Composition of Two Simple Harmonic Motions of the Same Period at Right Angles.—If the amplitudes of the two vibrations to be compounded are r_1 and r_2 respectively, construct a rectangle $BADE$ with sides $2r_1$ and $2r_2$ and two circles of diameters $2r_1$ and $2r_2$, as shown in Fig. 143. If C_1 and C_2 be points moving with the same uniform angular motion around these circles respectively, then the projection of C_1 on the horizontal or x axis will give one of the motions to be compounded, and the projection of C_2 on the vertical or y axis will give the other. The resultant of these two simple harmonic motions may then be represented by the motion of a particle whose equilibrium position is O and whose displacement from the equilibrium position is always inside the rectangle $BADE$.

The path or orbit of the resultant motion will depend on the phase difference between the two motions to be compounded. Suppose the x component is in advance of the y component by one-eighth of a period, that is $\delta = \frac{1}{8}T$, or the phase angle $e = 45°$. Suppose initially, that is when $t = 0$, the angle e and the points C_1 and C_2 are in the positions represented in Fig 143. This represents the motions given by the equations

$$x = r_1 \sin (\omega t + \pi/4), \; y = r_2 \sin (\omega t)$$

Divide each circle into the same number of equal parts, beginning at the initial positions of C_1 and C_2 and numbering these in regular order. It is evident that the successive positions of the particle

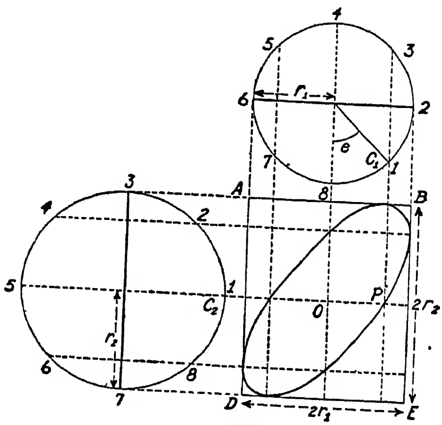


FIG. 143.

whose motion is that of the resultant of the two motions will be at the intersections of the lines 1-1, 2-2, 3-3, etc., and that a smooth curve drawn through these points will give the path of the resultant motion. For the case plotted the path is an ellipse inclined to the axis. If the phase difference is zero the path is a straight line, the diagonal BD. If $e = 90°$, or $\delta = \frac{1}{4}T$, the path is an ellipse with vertical and horizontal axes; this ellipse becomes a circle if $r_1 = r_2$. Paths or orbits for certain values of e from 0 to π are shown in the top row of Fig. 144, the sine functions being assumed for all cases. Other similar orbits would be obtained for values of e between π and 2π.

If the periods differ slightly, one vibration will gain on the other in phase, and the orbit will run through the complete cycle of forms

corresponding to $\delta = 0$ to $\delta = T$. The first half of these is shown in the top row of Fig. 144. If n_1, n_2 are the respective frequencies, the cycle will repeat itself whenever one component gains a whole vibration on the other or $(n_1 - n_2)$ times a second.

244. Composition of Two Simple Harmonic Motions at Right Angles with Periods in Simple Ratio.—Proceed as in the last

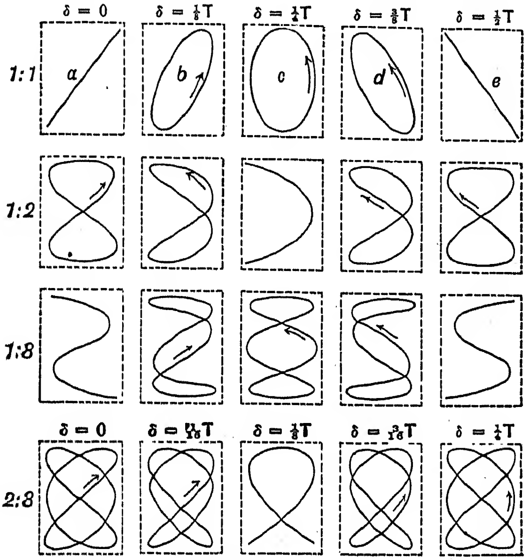


FIG. 144.

case, but divide the respective circles of reference into a number of equal parts, the number being proportional to the respective periods, so that the intervals in the two circles will be traversed in equal times. It will be noticed that in this case **the phase difference between the two components will be continuously changing,** whereas in the case where the two periods are the same the phase difference remains constant. Fig. 145 illustrates the case where $T_1/T_2 = 1:2$, the two kinematical equations being

$$x = r_1 \sin 2\omega t$$
$$y = r_2 \sin (\omega t + \pi/4).$$

These equations might equally well be written

$$x = r_1 \cos (2\omega t - \pi/2)$$
$$y = r_2 \cos (\omega t - \pi/4).$$

Fig. 144 illustrates a number of paths for which the y period T_2 is greater than the x period T_1 in the ratio stated. The phase difference (δ) is given as the fraction of the x period T_1 that x is ahead of y at time $t = 0$, assuming both x and y to be expressed as sine functions, and y to have zero phase.

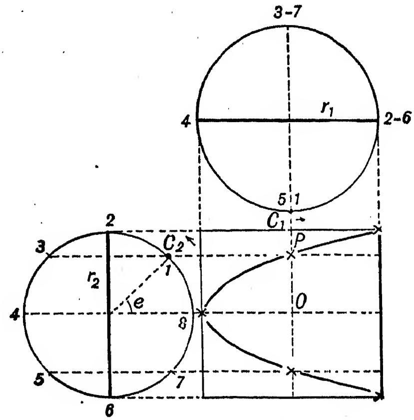


FIG. 145.

Experimental illustrations of these curves were obtained by Lissajous in the following manner. A beam of light is reflected from a mirror attached to one end of a tuning fork to a corresponding mirror on another fork vibrating in a plane at right angles to the first, and thence to a screen. The beam is displaced by both forks, and the spot of light on the screen describes the resultant path. Another method is to use a Y-pendulum, as shown in Fig. 146. If the bob vibrates in the plane of the paper, the effective length is PQ; if it vibrates at right angles to this plane, it is CQ. The periods in the two planes will, therefore, be different and independent. By properly adjusting the lengths PQ and CQ the bob may be made to describe the various Lissajous figures.

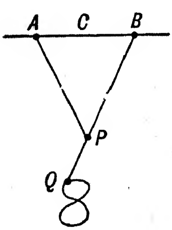


FIG. 146.

If the pendulum has a single support, $T_1 = T_2$, and the bob will move in an ellipse, circle, or straight line, according to the difference of phase between two impulses given to it at right angles.

A long, thin rectangular rod fastened at one end will vibrate transversely with a period depending on its thickness. If the diameters parallel to two sides are different, the respective periods

of vibration will be inversely as the diameters. If drawn aside diagonally and released, the rod will not continue to vibrate in that direction, but the displacement will be resolved into two components parallel to the diameters. If the ratio of the periods is simple, the end of the rod will describe Lissajous figures.

245. Waves Due to Simple Harmonic Motion.—Consider a number of spheres of equal masses attached to each other by elastic connections, as in Fig. 147. If a transverse simple harmonic vibration is imparted to the first, the impulse will be transmitted to the others in succession. Suppose the phase difference between the displacements of successive spheres to be one-eighth of a period. When *a* has completed one vibration, *b* has completed seven-eighths of a vibration, etc., while *i* is just beginning to move. The positions of the spheres will be at the projections on the

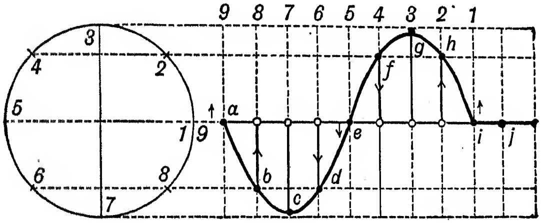


FIG. 147.

vertical lines 1, 2, 3, etc., of the points 1, 2, 3, etc., of a circle of reference, with radius equal to the amplitude of the wave. If a smooth curve be drawn through these positions, it will give the wave form. It is evident that the abscissa of any point on this curve is proportional to the time required for the disturbance to reach that point, or to the phase angle, and its ordinate to the sine of the phase angle of the disturbance at the point. Such a locus is called a **harmonic curve or sine curve**, and gives the shape of a transverse wave when the medium executes simple harmonic vibrations. If the particles in Fig. 138 execute simple harmonic vibrations, the longitudinal wave will be of the same type, and may be represented by a sine curve.

The **period** and the **amplitude** of the wave are the same as those of the simple harmonic motion of any point in the medium. The **wave-length λ is the distance between any two consecutive points in the same phase of displacement,** for example *a* and *i* (Fig. 147). If v is the velocity of propagation of the wave, $vT = \lambda$,

since λ is the distance transversed by the wave during a complete vibration of the "source," sphere a. If $n = 1/T$ is the **frequency** of vibration, $v = n\lambda$, the length of the train of waves sent out in one second.

The displacement in a longitudinal wave presents the same aspect if looked at from any direction in a plane at right angles to the direction of propagation This is not the case with the transverse waves represented in Figs. 137, 147, for the vibrations will be in the line of sight if viewed in the plane of vibration, and at right angles to the line of sight if viewed normally to this plane. These transverse waves have a sort of polarity, therefore, and are said to be *plane polarized*.

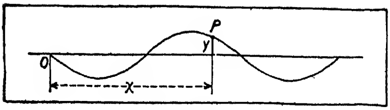


FIG 148

Transverse waves may be set up in a cord or longitudinal waves in a spiral spring by fixing one end and attaching the other to a vibrating tuning fork. The amplitude of the waves in such cases may be much greater than that of the fork.

If a beam of light be reflected from a mirror attached to the end of a vibrating fork, and again reflected to a screen from a revolving mirror, the harmonic curve will be traced on the screen by the spot of light Persistence of vision will cause the path to appear continuous.

A permanent record of such curves may be made by causing a bristle attached to the end of a tuning fork to trace its path on the smoked surface of a piece of glass which is moved past the fork at a uniform rate v.

246. Wave Equation.—To obtain the equation of a wave moving *forward* in the x-direction, we consider the relations in Fig 148. At the origin O a particle is vibrating in the y direction, just as in Fig 147, according to the equation

$$y = r \sin \omega t = r \sin 2\pi\frac{t}{T} \tag{1}$$

Each particle to the right of O, such as P, is vibrating with the same period T as the particle at O, but is *behind* in phase by an amount proportional to its distance x from the origin If this distance is just one wave-length λ, the phase retardation is just 2π If the distance, in fractions of a wave-length, is x/λ, the phase retardation is $2\pi(x/\lambda)$. Hence the general expression for the displacement y of a particle at any position x at any time t, is

$$y = r \sin (\omega t - \theta) = r \sin \left(2\pi\frac{t}{T} - 2\pi x/\lambda\right) = r \sin 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right) \tag{2}$$

This is the equation of a sine curve, repeating itself in intervals of time T, and in intervals of space λ. For a wave moving backwards along x, one should change the minus sign to plus, in this equation. Remembering that $\lambda = vT$, Eq. (2) can also be written

$$y = r \sin \frac{2\pi}{T}\left(t - \frac{x}{v}\right), \text{ or } y = r \sin \frac{2\pi}{\lambda}(vt - x) \tag{3}$$

The instantaneous "picture" of the wave, at a certain time t_0, is obtained by putting $t = t_0$ into equations (2) or (3). In particular, at $t = 0$ we get

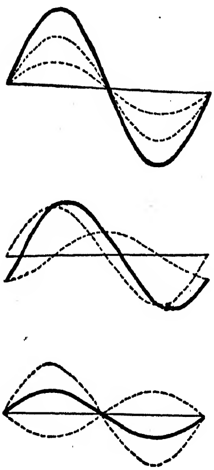
$$y = -r \sin \frac{2\pi x}{\lambda} \tag{4}$$

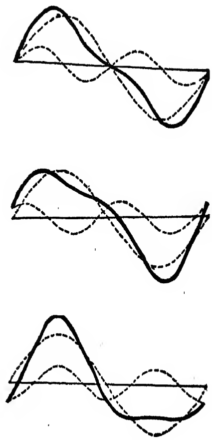
This is the equation of the curve drawn in Figs. 147 and 148. At any given point x_0, the motion is simple harmonic, with initial phase $\theta = 2\pi x_0/\lambda$. In particular, at $x = 0$, the motion is given by equation (1).

247. Superposition and Interference of Waves.—If two or more trains of waves are superimposed, each will give rise to independent displacements of the medium. The resultant effect may be obtained, therefore, by plotting each train of waves on the same axis, with relative displacements corresponding to their phase differences, and adding the ordinates. It is convenient to express the phase differences in terms of wave-length. If, for example, one wave starts half a period later than another, its front should be plotted half a wave-length behind that of the first. In Fig. 149, A, B, C, the full line represents the resultant of two waves of the same length and with phase differences of 0, $\lambda/4$, and $\lambda/2$, respectively. In the last case the resultant effect is zero if the amplitudes are equal. The modification of amplitude due to the superposition of waves is called **interference.** It is evident that the length of the resultant wave is the same as that of its components, and that it is a harmonic curve if they are harmonic curves.

248. Complex Waves.—Waves of different lengths, if moving with the same velocity, may be combined in the same manner. If one wave-length, λ_1, is $\frac{1}{2}$, or $\frac{1}{3}$, or $\frac{1}{4}$, etc. that of the other, λ_2, then one obtains a train of identically shaped waves of wave-length λ_2, but with a form depending on the relative wave-lengths, amplitudes, and phase relations. Thus in Fig. 150, $\lambda_1 = \lambda_2/2$, and the amplitude $r_1 = r_2/3$ for all three curves A, B, and C, but the phase relations are different. Fig. 151 illustrates the case where $\lambda_1 = \lambda_2/3$, $r_1 = r_2/3$, and the phase difference is zero at the starting point.

If the wave-lengths have a simple ratio such as $\lambda_1 = 5$, $\lambda_2 = 3$, the resulting disturbance consists of a repeating form of length $3\lambda_1 = 5\lambda_2$, since three waves of length λ_1 equal five of length λ_2. Fig. 152 represents the case where $4\lambda_1 = 3\lambda_2$, with equal amplitudes and starting in the same phase. Notice the region of small





A

B

C

Fig. 149.

Fig. 150.

displacements near the half-way point. This is due to the fact that the two wave trains are in the *same* phase at the two ends of the figure, but in opposite phase at the center, and in approximately opposite phase in the vicinity of the center. In general, if the frequencies of two wave trains are n_1 and n_2, there will occur, in the distance v (velocity of the waves), $n_1 - n_2$ regions of small displacements where the two trains are in approximately opposite phase, lying between an equal number of regions of large displacements. This is the graphical representation of "beat" waves in sound, described in §578.

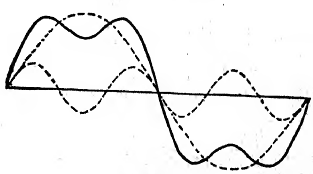


Fig. 151.

Such forms may be obtained experimentally by the optical method for obtaining sine curves described in §245, the beam of light being reflected successively from two forks vibrating in the same plane and giving beats, and from a rotating mirror to a screen.

The displacements of the medium may be the resultant of two displacements at right angles. The example of water waves has

already been mentioned. If one end of a cord be attached to the end of a rectangular rod vibrating transversely, so that the cord is parallel to the rod, the end of the rod will describe Lissajous figures (§244) and each element of the cord will do the same. If the two diameters of the rod are equal, each element of the cord will move in a circle or ellipse in a plane transverse to its length, but the phases will differ from point to point, so that at a given instant the cord will have the shape of a corkscrew. Such a wave is said to be *circularly* or *elliptically polarized.*

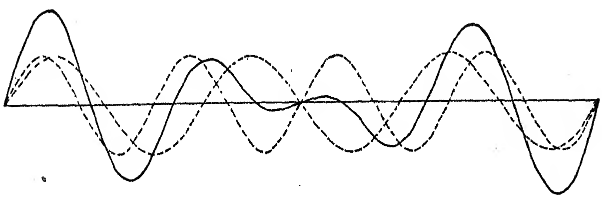


FIG. 152.

249. Fourier's Theorem.—The illustrations given show that various complicated forms may be obtained by the addition of simple harmonic waves of different lengths and phases, and that these waves will be of persistent form if the periods of the components are simple fractions of the longest component. Fourier proved that any periodic disturbance or wave form of permanent type could be represented as the summation of a number of simple harmonic terms of the form

$$x = r_1 \sin \omega t + r_2 \sin 2\omega t + r_3 \sin 3\omega t + \ . \ . \ . \ , \text{etc.}$$

the periods and wave-lengths of the components having the ratios 1, $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, etc. Fig. 151 shows that the resultant is approaching a rectangular form which may be more nearly attained by adding shorter waves.

The forms of complex waves may be projected by the following device (Fig. 153): A screen with a vertical slit opening at O is placed in front of a horizontal stretched wire AB, which is illuminated from the lens L. A sharp shadow of the segment opposite the opening may be thrown on a screen S after reflection from a rotating mirror M. If the wire is at rest the shadow of the segment will be drawn out in a dark straight line on the screen. If the wire vibrates in a vertical plane the shadows of the segment in its successive phases as the wave passes O will be laid off end to end on the screen, giving the actual form of the wave passing the opening.

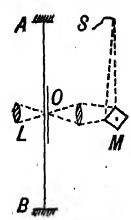


FIG. 153.

250. Velocity of a Wave on a Cord.—Let the wave be supposed to be moving toward the left with a velocity v. It will simplify the problem without essentially changing it, if we now suppose

the cord to be given a velocity v toward the right. The wave will then stand still and every part of the cord, as it comes to the wave, will pass through it with velocity v. This, in fact, is what may often be noticed in the use of a chain hoist. If the chain be started in rapid motion (there being no load on the lower pulley), a bend impressed on the chain will sometimes remain stationary for a short time, and, if the chain be suddenly arrested, the bend will move off in the opposite direction with (approximately) the speed which the chain had. The *relative* velocity depends only on the mass and tension of the chain.

Now let QR be a small part of the wave, its length l being so short that it may be regarded as an arc of a circle (the circle of curvature). Draw tangents TQ and TR and complete the parallelogram $QTRS$. The velocities at Q and R may be represented

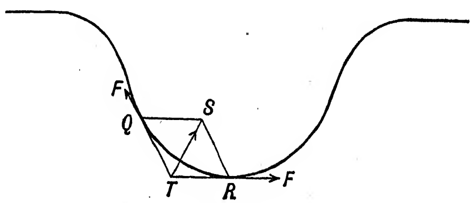


FIG. 154.

by QT and TR. As each part of the cord passes from Q to R in time t it will have an acceleration, a, toward the center of curvature, that is, in the direction of the diagonal TS. Since TS represents a velocity which, added to QT, gives QS or TR, it represents the change of veloci,ty at, in the time t. Hence

$$\frac{TS}{QT} = \frac{at}{v}$$

The only forces that act on the part QR of the cord are the equal forces, F, at its ends due to the tension in the cord. By the proper choice of scale these also may be represented by TQ and TR and their resultant, represented by TS, is the force that causes the central acceleration of the part QR of the cord. If the mass per unit length of the cord is m, the mass of the part QR is ml. Hence

$$\frac{TS}{TQ} = \frac{mla}{F}$$

Since in these equations only numerical magnitudes are represented,

$TQ = QT$. Hence, equating these values of TS/QT and noting that $l = vt$, we get

$$v = \sqrt{\frac{F}{m}}$$

(It is evident that belting traveling with this velocity will exert no pressure on a pulley. See §49.)

251. Velocity of Elastic Waves.—It might be expected that the velocity of waves in an elastic medium would depend upon the elasticity and density, since the elasticity determines the rate at which an impulse is transmitted from one element to another (in a perfectly rigid and incompressible medium the effect would be instantaneous), and since the density exercises a retarding influence, on account of the inertia of the displaced elements. The exact relation between the velocity, the density ρ, and the coefficient of elasticity E may easily be found in some cases.

For example, suppose the front of the disturbance in a longitudinal wave in a medium of unit cross-section is at A (Fig. 155) at one instant and at B a short time t later. The velocity of the wave is, therefore, $v = l/t$, where $l = AB$. An imaginary plane A in the medium is displaced to D, a distance x, by compression. If l is a very small fraction of the wave-length, the density of the substance will be practically uniform between D and B, and the center of mass of the element is displaced from C to C', a distance $x/2$. The average velocity of the center of mass is $x/2t$ and its final velocity x/t. The final force acting on the element is Ex/l, where E is the appropriate modulus of elasticity (see below). The average force is half of the above. Equating the work done by this force to the acquired kinetic energy due to the motion of the center of mass, we have

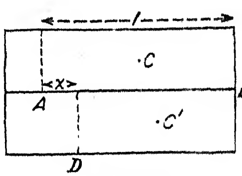


FIG. 155.

$$\frac{Ex}{2l}x = \tfrac{1}{2}\rho l\frac{x^2}{t^2};\ \text{therefore}\ \sqrt{\frac{E}{\rho}} = \frac{l}{t} = v, \text{ the velocity of the wave.}$$

This equation applies to longitudinal waves, such as sound waves, in any fluid, where E is the bulk modulus k, and to the so-called longitudinal waves (actually accompanied by lateral shrinkage and expansion) in a thin wire or rod, where E is Young's Modulus M.

252. Reflection of Waves.—When a transverse wave reaches the fixed end of a cord, the displacement is immediately reversed in direction by the elastic reaction of the fixed end. The wave is, therefore, reflected with reversal of phase of displacement. Appar-

ently the incident wave has disappeared through the end, while a wave of opposite displacement has entered, as shown in Fig. 156, traveling in the opposite direction, and at every instant exactly neutralizing the displacement of the end which would be caused by the incident wave if the end were free. When a continuous train is reflected, the effect is as though a train of indefinite length had been cut in two when a wave-front reaches A, the fixed point (Fig. 156), and the waves to the right immediately reversed in direction, while the incident waves continue their motion unchanged; or as though a train of incident waves were traveling through a mirror, while their **inverted** images proceed out of it in the opposite direction.

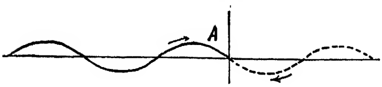


FIG. 156.

If one end of the cord is free, when the wave reaches that point, the end, having nothing beyond to restrain it, has an outward displacement twice as great as though the cord were continuous, and it will, therefore, immediately start a wave of the same phase in the reverse direction. After half a period of vibration it will return through the resting point in the opposite direction, and will start a backward wave with phase opposite to that of the incident wave. It is as though a train has been cut in two when a crest is at the free end B (Fig. 157), and the right hand section immediately reversed in direction; or as if an advancing train were passing through a mirror while its **erect** image emerged from it.

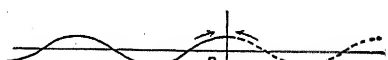


FIG. 157.

To show in another way the difference between reflection at a free end and at a fixed end, suppose that the part of the direct wave-train that first reached B and begins to be reflected is as represented in Fig. 158. If we compare this with Fig. 156, it is seen that in reflection at a free end, as compared with that at a fixed end, *there is a delay of half a period* in the reflection of the wave of opposite phase, as though the right-hand section in Fig. 156 were held at rest for half a period before starting in the negative direction.

FIG. 158.

Reflection of longitudinal waves may be illustrated by the conduct of a row of elastic pendulums of the same size, as shown in Fig. 159a, the last resting against a fixed obstacle. If a is drawn

aside and released, it will impart an impulse to b, this in turn to c, etc., and a compression wave will travel to the other end of the row; g cannot move, but will be compressed, and through its elastic reaction it will almost immediately start a compression wave in the opposite direction. When this wave reaches the free end, a will fly out without restraint, leaving a rarefaction behind it; or, if elastically connected with b, it will at once send back a rarefaction wave. In any event, after executing half a vibration it will swing back through its equilibrium position and reflect a compression wave to the right.

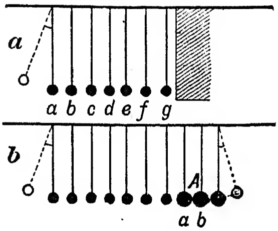


FIG. 159.

If there are two rows of elastic pendulums of different masses, as in Fig. 159b, displacements will be immediately transmitted across A, no matter in which direction the wave is moving, but the wave will also be partially reflected. If the wave travels to the right, reflection of a at A will be almost immediate; if it travels to the left, the more massive sphere b will continue after impact to move to the left, and will return through its resting point, to send a wave back to the right, at the expiration of half its period of vibration.

The two cases of reflection we have now considered illustrate the general principle that *the displacements in a medium have a minimum amplitude at a fixed or constrained boundary; a maximum amplitude at a free boundary or one with diminished constraint.* Important illustrations of this principle occur in cases where waves pass from a light to a dense medium or *vice versa* (§§585, 673).

253. Stationary Waves.—Consider a train of waves in a cord moving to the right, while a similar train (reflected or independent) moves to the left. Interference will take place, and the resultant displacement of the medium at a given point and time will be the sum of the individual displacements. Plot the positions of the waves at successive instants (say at intervals of an eighth of a period). If the incident train is represented by a light line, the reflected train by a dotted line, and the resultant by a heavy line (Fig. 160), it will be seen that there are always points of zero displacement N (or of minimum displacement if the amplitudes are unequal) at intervals of half a wave-length, where the waves always meet in opposite phases. Half way between these points, at L,

the waves will always meet in the same phase, and the displacement will be a maximum. The former positions are called **nodes,** the latter are called **antinodes.** Between the nodes the medium oscillates back and forth, the direction of the displacements being opposite in adjacent segments, so that at any instant the cord has a more or less sinuous shape, except at intervals of half a period, when it passes through the undisturbed straight position (center drawing of Fig. 160). The same conclusions apply to longitudinal waves. Disturbances of this sort are called **stationary** waves. It is evident that when these arise from the interference of incident and reflected waves there must be a node at a fixed or constrained boundary, an antinode at a (relatively) free or unconstrained boundary.

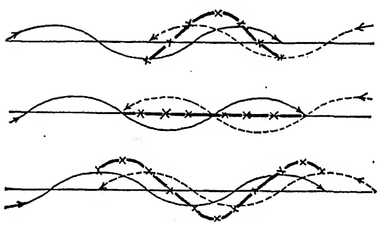


FIG. 160.

Fig. 161 is the graphical representation of stationary waves of longitudinal type. The displacements have just begun to return from maximum elongation, from the full to the dotted line. This indicates that the particles to the left of N_1 and those to the right of N_2 are moving in the negative direction, while those between N_1 and N_2 are moving in the positive direction. Consequently the particles on opposite sides of N_2 are approaching that point, while those on opposite sides of N_1 are receding from it. At N_2 there will be a condensation, at N_1 a rarefaction. After half a period conditions will be reversed. In the neighborhood of L_1, however, the particles are moving in the same direction with approximately the same velocity, so that their relative positions are only slightly changed. This is true also for L_2. It follows that **at the nodes there are the greatest variations of pressure and the least motion, at the antinodes the smallest variations of pressure and the greatest motion.**

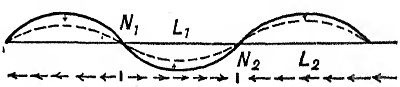


FIG. 161.

254. Waves on a Liquid.—Some of the most interesting properties of wave motion may be illustrated by waves on the surface of a liquid, such as water. The initial displacement may arise from differences of level caused by some external force, for example the

impact of a pebble, winds, etc. The effect of gravity, of fluid pressure, and of surface tension is to restore the original level, but, on account of their inertia, the particles are displaced beyond their equilibrium positions, just as in the case of vibrations of a liquid in a U-tube. Horizontal as well as vertical displacements must occur, as in the case of the liquid in the bend of the U-tube. There is, therefore, a longitudinal as well as a transverse component. These displacements are simple harmonic, because the resultant pressure on an element is proportional to its vertical displacement from the undisturbed surface. We have seen (§236) that on a crest the element moves forward, in the hollow backward, in intermediate positions both vertically and horizontally. Fig. 162 shows the positions and directions of rotation of a number of particles originally at rest on the surface in the positions under *a*, *b*, *c*, etc., the phase difference between successive displacements being an eighth of a period. Particle *a* is subject solely to a down-

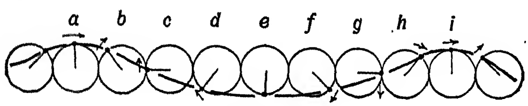


FIG. 162.

ward acceleration, particle *e* to an upward acceleration; particles *c* and *g* are subject solely to horizontal accelerations, due to the lateral pressure, as they are in the horizontal plane of equilibrium. We thus find that there is a difference of phase of a quarter period between the vertical and horizontal accelerations, in accordance with the observed fact that the disturbed elements move in circular or elliptic orbits. It is evident that *the wave form is not a sine curve.*

The expression for the velocity of waves on the surface of a liquid is complicated and cannot be derived here. It is sufficient to say that large waves are maintained by gravity alone, and that the velocity is independent of the density of the liquid, since the force acting is proportional to the weight of the displaced elements, and hence will produce the same acceleration, whatever the density. The velocity increases with the wave-length, so that one may frequently see a train of long water waves sweeping through a train of shorter waves and leaving them behind. When the liquid is shallow, the velocity diminishes with the depth. On the other hand very small waves are maintained by surface tension alone, so that

they are analogous to transverse waves in an elastic membrane. In the case of these waves the velocity increases as the wave-length diminishes, and is also dependent upon the density and the surface tension. Such waves are called **ripples.**

The complex general expression for the velocity of large water waves becomes simple for two extreme cases, which are of great interest. First consider ocean waves, when the wave-length is small compared with the depth. In this case, if the wave-length is λ and the velocity is v,

$$v = \sqrt{\frac{g\lambda}{2\pi}}$$

Thus the velocity is proportional to the square root of the wave-length and independent of the depth.

Next consider long waves in a canal, where the wave-length is large compared with the depth h. In this case approximately

$$v = \sqrt{gh}$$

so that the velocity depends only on the depth.

Another important consideration arises when we have, not a single ocean wave, but a group of waves, started by a disturbance or by the wind. Individual waves of different wave-lengths in the group travel with different velocities. This is called *dispersion.* Usually the wave-lengths differ only sightly from a mean value, λ, and then the group, while continually dissolving and reforming, is a complex pattern with a definite velocity of its own. The relation between the ***group-velocity*** u and the ***wave-velocity*** v corresponding to the wave of mean wave-length λ is, for these water waves, remarkably simple:

$$u = \frac{1}{2}v = \frac{1}{2}\sqrt{\frac{g\lambda}{2\pi}}$$

In the case of canal waves there is no dispersion, since the velocity is independent of the wave-length. These considerations regarding dispersion or its absence apply also to radiation and the wave theory of matter (§758).

So-called "tidal" waves, due to sub-oceanic earthquakes and affecting a whole ocean, are of the nature of waves of enormous wave-length, and their observed velocities give a rough estimate of the mean depth of an ocean (see problem 11).

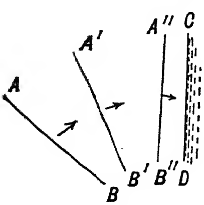


Fig. 163.

255. Refraction of Waves.—Waves move more slowly in shallow than in deep water. Hence if the front AB of an ocean wave moving in the direction of the arrow (Fig. 163) approaches a beach CD, the nearer end, B, of the wave will be retarded more than A, being in shallower water. The wave front will swing around into the successive positions $A'B'$ and $A''B''$, and will finally become approximately parallel to the shore line. This **change in**

direction due to change in velocity is called **refraction.** Similar effects are, we shall see, shown by other waves, such as sound and light, when they pass from one medium to another in which they travel with different velocity.

256. Propagation and Reflection of Ripples.—Experiments with ripple waves may be shown by the following arrangement. A shallow wooden box with a glass bottom, about two feet square, is mounted on legs like a table, carefully leveled, and partly filled with water. Light may be projected upward through the bottom from an arc placed beneath the box, or by reflecting a divergent beam of sunlight upward by an inclined mirror. Ripples on the surface will by their lens effect change the distribution of light on the ceiling so that the motion of each ripple may be followed.

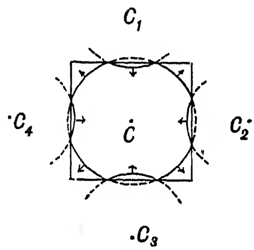


FIG. 164.

If the middle of the surface is touched with a nail, a circular ripple will diverge from that point. If the surface were larger, this wave would at a later time occupy the position of the circle (Fig. 164), but it will be in part reflected from the four sides. The reflected segments are exactly like the missing segments of the outgoing wave, reversed in direction. These reflected waves have centers at C_1, C_2, C_3 and C_4, the "images" of the source C, which are evidently at the same distance from the walls as the source itself, since C and the other centers of curvature are symmetrically situated with respect to the walls. These reflected waves will cross each other and be subject to repeated reflections ("multiple reflection"), their curvature all the while decreasing, until we have a rectangular system of straight ripples.

If a circular wave strikes a bent sheet of metal of the same curvature as the wave, the latter will be reflected without change of curvature, converge to its starting point, and diverge from it on the opposite side (Fig. 165*a*). If the strip has a greater curvature than the wave, the edges of the latter will be first reflected, so that its curvature is increased (*b*). It will converge to C', a "real image" or "focus." If the strip is concave with less curvature than the wave (*c*) the latter may be made divergent, with a "virtual" image at C'. If the strip is convex toward the wave, the reflected wave will always diverge from a virtual center behind the mirror (*d*).

If the surface is touched with a long straight strip of metal a straight ripple will be produced. If this strikes a screen with a small slit in it (Fig. 166*e*) the disturbance will pass through this hole and set up a semicircular wave on the other side. The remainder of the wave will be reflected as a straight line.

If a number of nails are driven at equal distances through a strip of wood and dipped into the water, circular waves will diverge from the points of contact. At a little distance these wavelets will blend into a straight ripple corresponding to their common tangent (*f*). At other points the ripples cross each other in all phases, and their effect will vanish because of interference. We may, therefore, consider that a linear wave front is due either to a continuous linear disturbance

or to a number of neighboring point disturbances, each sending out circular waves. In (*e*) for example only the point in the opening is effective for transmission. The latter conception is often useful (§627).

If a screen *S* projects part way across the tank (*g*), the portion *AS* of an incident wave will be reflected; the remainder *SB* of the wave will pass the screen. It will be noted that the end of the transmitted wave front will bend into the shadow of the screen, and the end of the reflected wave will bend into the region formerly occupied exclusively by the other half of the wave. *S* is apparently a center of disturbance for both these waves. This effect is called **diffraction.** By noting the resemblance of the ends of the waves in this case to those in the preceding case (*f*) the explanation will be made clear.

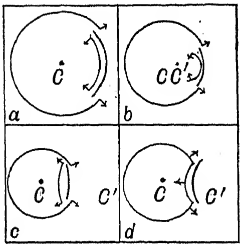


FIG. 165.

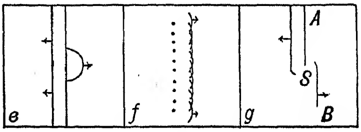


FIG. 166.

257. Refraction of Ripples.—Advantage may be taken of the fact that the velocity of water waves diminishes with the depth to illustrate refraction. On the bottom of the tank lay a piece of thick glass (Fig. 167), so that the water over it is about one-fourth as deep as elsewhere. If a circular wave is started by touching the surface at *C*, the middle will be more retarded than the edges when the wave comes from the deeper water, and the curvature of the wave will be diminished (*h*). If the wave travels from the shallow regions, the contrary will be the case (*i*). The centers of curvature or "images" of the source will be at *C'* (outside the tank in *h*).

If a prismatic sheet of glass is laid on the bottom (*j*) a linear wave front *AB* will be rotated both in approaching and leaving, and the final direction will be *A''B''*. If pieces of glass with convex or with concave edges, like sections of lenses, are laid on the bottom, the center of a passing circular wave will be more retarded than the edges in the first case (*k*), and less retarded in the second (*l*), resulting in changes of curvature. The "images" of the source will be at *C'*.

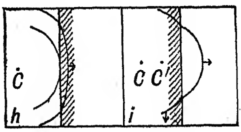


FIG. 167.

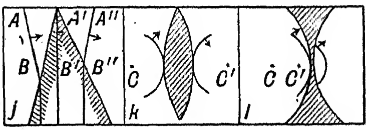


FIG. 168.

258. Interference of Ripples.—If two nails simultaneously touch the water at different points two circular waves will be set up, which will cross and interfere with each other. They pass so quickly, however, that it is difficult to observe them. Better results will be secured if a continuous series of waves can be produced, and still better results if there is a system of stationary waves. A very

satisfactory method of securing this result is to put mercury in a circular glass dish at least four inches in diameter, and maintain periodic disturbances at the center by a glass fiber attached to the vibrating prong of a tuning fork. Continuous trains of circular ripples will diverge from the center, while reflected circular ripples will converge toward that point. The result will be a system of circular stationary waves, as illustrated in Fig. 169. They may be projected on a screen by reflected light, and made more distinct by using a lens.

If two glass fibers are attached to the fork near each other, two trains of waves will be maintained, and each will form its own system of stationary waves. At all points on the surface where the outgoing waves meet each other in the same phase (that is, where the difference of the respective distances to the two sources is zero or any whole number of wave-lengths) the waves will reinforce

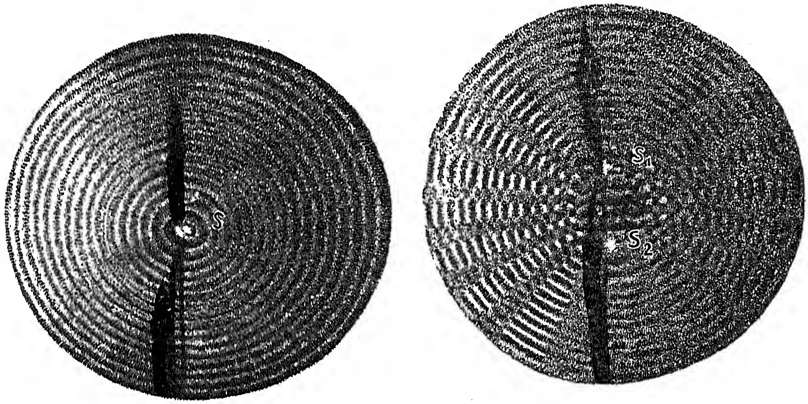


FIG. 169.

each other. In regions where they meet in opposite phases (the differences of path being some odd multiple of a half wave-length), they will destructively interfere with each other. Along certain lines, therefore, there will be no disturbance by either outgoing or reflected ripples (Fig. 513). Between these lines segments of the stationary waves appear, as shown in Fig. 169.

259. Energy and Intensity of Waves.—The energy of a vibrating body is proportional to the square of its amplitude (§64). Each vibrating element of mass in a medium traversed by waves will, therefore, possess energy proportional to the square of its amplitude, and this energy will flow forward with the advancing waves. The **intensity** of waves in a given region is defined as being proportional to **the amount of energy passing per second through unit area at right angles to the direction of propagation,** hence the intensity is proportional jointly to the square of the amplitude and to the velocity of the waves. In a viscous medium, such as

molasses or lead, the waves rapidly decay in amplitude and disappear, owing to the absorption of energy by internal friction. This effect is known as **damping.** Fig. 139 represents the form of a damped train of waves. If there is no such loss the same quantity of energy will persist in a given wave, no matter how far it travels, or how the dimensions and form of the wave front may change. If such waves travel in a wire or any other channel of constant cross-section the intensity will be independent of the distance from the source, as the wave front will remain of constant area. This is illustrated by the transmission of sound waves through a speaking tube or of light waves in a parallel beam. In the case of circular waves on a surface, a constant amount of energy will remain in a wave of circumference which increases directly as the distance from the source; hence the intensity must vary inversely as the distance, and the amplitude inversely as the square root of the distance. In the case of spherical waves, the energy will remain constant within a spherical shell of the thickness of one wave length and with surface increasing as the square of the distance. If F is the energy emitted from the source per second, and if r_1 and r_2 are the radii of the wave at different distances, and E_1 and E_2 the corresponding intensities, by equating two expressions for the energy emitted from the source per second we get

$$F = 4\pi r_1{}^2 E_1 = 4\pi r_2{}^2 E_2 \therefore \frac{E_1}{E_2} = \frac{r_2{}^2}{r_1{}^2}$$

Hence the intensity varies inversely as the square of the distance from the source, and the amplitude inversely as the distance.

PROBLEMS

1. A mass of 980 grams is suspended on a spring of negligible mass and of such stiffness that a force of 50 grams weight will stretch it 2 cm. The mass is pulled 5 cm. below the point of equilibrium and released. Find (*a*) the period; (*b*) the kinematical equation, giving the motion of the center of mass, counting time from the instant of release and using the cosine form; (*c*) the displacement at time = 3 sec.; (*d*) the unbalanced force acting at this time; (*e*) the velocity at time = 2 sec.; (*f*) the maximum kinetic energy.

Ans. (*a*) $T = 0.4\pi$ sec.
(*b*) $y = 5 \cos (5t + \pi)$.
(*c*) $y = +3.773$ cm.
(*d*) $F = -92\,450$ dynes.
(*e*) $v = -13.616$ cm./sec.
(*f*) $K.E. = 306{,}250$ ergs.

2. Water or mercury in a U-tube is disturbed. Show that the liquid executes a simple harmonic motion of period $T = 2\pi\sqrt{l/2g}$, where l is the length of liquid from surface to surface around the bend.
3. Two simple harmonic motions in the same line are given by

$$y_1 = 4 \cos 3t \qquad y_2 = 3 \cos (3t - 3\pi/4)$$

Find the amplitude and initial phase of the resultant motion. What is the equation of the resultant motion?

Ans. $R = 2.8336$
$\theta = -48°28'$
$Y = 2.8336 \cos (3t - 48°28')$

4. Compound graphically three trains of waves of lengths in ratios 1, $\frac{1}{3}$, and $\frac{1}{5}$ and of amplitudes 3, 2, and 1, starting in the same phase.
5. Compound graphically two trains of waves of lengths 5 and 4 and of equal amplitudes.
6. The velocity of sound in air-free water at 19° C. is given by the Smithsonian Tables as 1461 meters per sec. What is the resulting modulus of elasticity of water at this temperature, in dynes per sq. cm.? What is the coefficient of compressibility per atmosphere?

Ans. $E = 2.1345 \times 10^{10}$ dynes/sq. cm.
$1/E = 4.743 \times 10^{-5}$ per atmosphere.

7. A copper wire two sq. mm. in cross-sectional area is subject to a tension of 5 kg. weight. With what velocity will a transverse wave travel along it?

Ans. 5.2408×10^3 cm./sec.

8. With what velocity will an essentially longitudinal wave travel in the same wire?

Ans. 3.5116×10^5 cm./sec.

9. A steel wire one mm. in radius is stretched with a tension F between two supports. At the same instant a longitudinal and a transverse wave are started from one end. At the time the longitudinal wave has arrived at the other end, the transverse wave has traveled only 1/100 of the length of the wire. Find the tension in kilograms weight.

Ans. 7.373 kg. wt.

10. The Suez canal is somewhat less than 32 ft. deep, but, taking it as 32 ft. and g as 32 ft./sec.^{2}, what is the velocity of a long wave in the canal?
11. An earthquake tidal wave, originating at Simoda, Japan, reaches San Diego 4920 miles away in 12 hr. 13 min. Calculate the mean depth of the ocean between the two places.

HEAT

BY CHARLES E. MENDENHALL, PH. D.

Late Professor of Physics in the University of Wisconsin

REVISED BY L. R. INGERSOLL, PH. D.

University of Wisconsin

INTRODUCTION

260. Early Ideas.—The preceding sections have dealt with physical changes involving, in general, motion and changes in motion of bodies as a whole. We have now, however, to consider changes in physical condition which do not involve obvious changes in motion, of which the most common are changes in hotness or coldness and changes in state, that is, melting or boiling. The sense of touch is the first and simplest means of distinguishing hot from cold bodies, and by it we can roughly arrange bodies in the order of their hotness, deciding that A is hotter than B, B than C, etc. But the sense of touch is found to be neither reliable nor delicate enough to be used as a *measure* of degrees of hotness, and, moreover, a limit of hotness or coldness is very soon reached, beyond which the touch sense cannot be directly applied. A purely physical basis of measurement (§263), depending on the properties of bodies, is therefore adopted, which agrees with the sense of hotness as far as they can be compared. When measured in this definite physical way, the hotness of a body is called its *temperature,* the scale of measurement being so chosen that hotter bodies have higher temperatures.

It is found that increase in temperature of a given body can be produced by various common causes, such as contact with or exposure to fire, contact with a hotter body, and friction, as, for example, rubbing one's hands together. The causes which will produce increase in temperature will also, under proper conditions, produce melting or boiling and various other physical changes, of which increase in size is the most common and obvious. On

account of these common causes, it was most natural to group together the various effects referred to as they became known and attribute them all to the passage, into or out of bodies, of a substance called *caloric* or *heat*, the presence or absence of which accounted for all of these related phenomena. According to this theory, heat was a material substance, but one which could not be weighed or detected by any ordinary physical method. On the basis of this hypothesis fairly consistent explanations were given for many common facts. For example, the temperature of a body was said to depend on the amount of caloric it contained and upon its natural capacity for caloric, which in turn depended upon its physical state, as, for instance, the state of subdivision. A given amount of matter in powdered form was thus supposed to have a less capacity for caloric than the same quantity in larger pieces. Thus the rise in temperature produced by rubbing two bodies together was explained as being due to the abrasion of the material, its capacity for caloric being thereby reduced and a certain proportion of its caloric set "free," and its temperature correspondingly raised. According to this idea, the entire amount of caloric set free should under given circumstances, be proportional to the entire amount of material abraded.

261. Heat and Work.—The first serious question of the truth of the caloric theory was raised in 1798 by Count Rumford, who, in experiments carried out in Munich upon the caloric developed in the boring of cannon, used a blunt borer which cut very little material, and arranged matters so that the heat generated raised the temperature of a considerable quantity of water, which was made to boil "without fire." From these experiments he concluded that the amount of caloric developed was not at all proportional to the amount of abrasion, but was, at least approximately, proportional to the amount of *mechanical work* required to do the abrading. In the following year Sir Humphrey Davy performed the similar but more striking experiment of melting ice by rubbing two blocks of it together, the temperature of the ice as a whole being below freezing. In this case the "abraded material" was *water*, which was well-known to have a *larger* caloric or heat capacity than ice, so the argument mentioned in §260 could not be used here. Accordingly it was again concluded that the melting was due to the transmission of *motion* to the ice molecules. From this time on, the idea that heat could be produced from mechanical

motion and *vice versa*, or, as it is put to-day, *that* **heat is a form of energy**, was gradually accepted. But it was nearly 50 years before the full significance of this new point of view was appreciated and careful measurements were made by Joule and others of the amount of work equal to a given amount of heat. This idea that heat is a form of energy, together with the ideas of the kinetic theory of gases (§227), and the conception of the molecular structure of matter, suggested by chemical and radioactive (§546) investigations, unite to give the present *molecular* or *kinetic theory of heat.*

262. Molecular Theory.—According to this point of view, matter consists of units or parts called *molecules*, which are composed of smaller units of the elements (oxygen, hydrogen, iron, etc.), called *atoms*, these in turn containing still smaller units, namely elementary charges of negative electricity called *electrons* (§366) and a *nucleus* or center, having effectively a positive electric charge, and probably built up of a combination of electrons and units of positive charge called *protons*, or, more likely, a combination of protons and neutral particles of proton mass, called *neutrons*. Our knowledge of the structure of atoms is still very imperfect, but the electrons in the atoms undoubtedly move about very considerably, probably somewhat as planets move about the sun, while the atoms move about inside the molecule, and molecules move inside the mass of matter, with great freedom when the matter is gaseous, with less freedom when it is liquid or solid (§§157–161). It is also possible under various conditions to have electrons existing more or less independent of atoms as "free" electrons or negative electric charges, the atoms which have lost electrons then having a positive electric charge and being ready to capture any other electron which happens to come near enough; free electrons are characteristic especially of metals. Broadly speaking, the addition of heat energy to a body either increases the (kinetic) energy of motion of its molecules or increases their (potential) energy of position, as when melting or boiling occurs.

Considering this more in detail we see that all of the possible motions of molecules, atoms, and electrons would involve kinetic energy. Moreover, it is evident that changes of position of molecules, atoms, and electrons with respect to each other, against the forces, probably entirely electrical in nature, existing between them, would require work against these forces, that is, changes in

potential energy. Hence we can see that, when heat energy is added to a body, it may appear:

1. As an increase in the kinetic energy of motion of the molecules and free electrons.

2. As an increase in the potential energy of the molecules with respect to each other, if their average distance apart is increased.

3. As an increase in kinetic and potential energy of atoms and electrons inside the molecules.

This analysis of the possible changes in what is called the *internal energy* of bodies should be kept in mind throughout the study of heat, which is largely a study of the effects of changes in internal energy upon the condition and properties of matter. To start the study we need a clear definition of temperature.

THERMOMETRY

263. Definition of Temperature.—For the most accurate work, temperature is measured on the *international thermodynamic scale* (§351); but, as the principles involved in the definition cannot be explained until later, we shall begin with a scale that was, until recently, the international standard and differs only *very* slightly from the present one. In terms of the pressure of hydrogen, *changes of temperature are defined as being proportional to the corresponding changes of pressure in a constant mass of hydrogen confined at constant volume.* This is called, *the hydrogen constant volume scale.* To measure the temperature of a body, for example a mass of water, the vessel containing the hydrogen would be held in the water and the pressure of the hydrogen measured. But before temperature can be expressed as a *number*, we must have a unit in which to express it and we must also agree on a reference point or "zero" from which it is to be measured. The ordinary zero, called the "ice-point," is the temperature of a mixture of pure ice and water when the pressure on the water surface is 1 atmosphere, and the degree is fixed by adopting a second standard point, the "steam-point," or the temperature of boiling water when the pressure is 1 atmosphere, which is specified as $+ 100°$ or $100°$ above zero. The *degree* is then such a change in temperature as will produce $\frac{1}{100}$ the change in pressure which is observed when the hydrogen is heated from the ice-point to the steam-point. These specifications

define the *Centigrade zero* and *Centigrade degree,* which are universally used in scientific work.

A *thermometer* is an instrument for measuring temperature according to some definite scale. A *constant volume gas thermometer* is an apparatus for measuring temperature by the variation in pressure of a gas confined at constant or nearly constant volume. If the gas used is hydrogen the thermometer gives at once standard temperature; with other gases it must be calibrated in terms of the standard. Such an arrangement is shown diagrammatically in Fig. 170, and consists essentially of a bulb of glass, glazed porcelain, fused quartz, platinum or platinum-iridium (according to the temperature range over which it is to be used), connected by a capillary tube to a mercury pressure-gauge, such as the open manometer shown. The pressure of the confined gas can be measured by reading the difference in level of the two mercury columns and adding to this the atmospheric pressure as determined by a barometer.

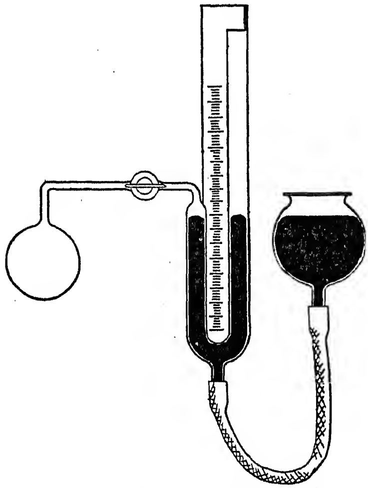


FIG. 170.—Constant volume gas thermometer.

Still keeping the pressure of hydrogen at constant volume as the basis of the temperature scale, other numbers may be assigned to given temperatures by giving another number to the ice-point and subdividing the interval from the ice-point to the steam-point into a different number of degrees. In this way the Fahrenheit scale (the one in ordinary use in English-speaking countries) is obtained by giving the value 32 to the ice-point and subdividing the interval from the ice-point to the steam-point, the *fundamental interval* as it is called, into 180°. (Fahrenheit originally used other temperatures to define his scale; a freezing mixture of water, ice, and salt gave what he called 0°, and blood heat what he called 96°.) From the above statements we derive the following conversion formula for changing the Fahrenheit to the Centigrade scale:

$$(t_F - 32)\frac{5}{9} = t_C$$

The choice of a *thermometric property* (in this case *pressure* of *hydrogen*) is entirely independent of the choice of numerical scale, *i.e.*, zero and size of degree; the Centigrade or Fahrenheit numerical scale can each be applied to any other thermometric property.

TABLE 1

CERTAIN TEMPERATURES ON THE CENTIGRADE SCALE, WITH THEIR (APPROXIMATE) EQUIVALENTS ON OTHER SCALES

Centigrade	Absolute (Kelvin)	Fahrenheit	Occurrence
−273.2°C.	0°K.	−459.8°F.	Absolute zero
<−273.1°	<0.1°	< −459.7°	Lowest measured temperature
−190°	83.2°	−310°	Liquid air under 1 atmosphere pressure
− 62.8°	210.4°	− 81°	Stratosphere at 22 km (1935 flight)
0°	273.2°	32°	Melting point of ice
57.2°	330.4°	135°	Highest recorded atmospheric temperature (Arabia)
100°	373.2°	212°	Boiling point of water under 1 atmosphere pressure
700°	973°	1292°	"Dull red" heat for solids
1600°	1873°	2912°	"White heat" for solids
2150°	2423°	3902°	Tungsten filament lamp (vacuum)
2450°	2723°	4442°	Tungsten filament lamp (gas-filled)
3600°	3873°	6512°	Carbon arc
6000°	6273°	10832°	Surface of sun (estimated)
>20000°	>20273°	>36032°	Highest artificially produced temperature (exploded wire)
100000°	100273°	180032°	Surface of hottest stars (estimated)

It is found that the change in pressure (volume being constant) of hydrogen for 1°C. (as defined in the preceding) is $\frac{1}{273.04}$ of the pressure at 0°C. Hence, if the same scale of temperature were carried below zero Centigrade (Fig. 171), the pressure would be reduced to zero at a temperature of −273.04°C. This is called the *absolute zero of the hydrogen constant volume scale,* and,

according to the ideas of the kinetic theory of gases (§227), it corresponds to a state of zero molecular velocity, since pressure is due to the impact of moving molecules. This temperature could not, however, be measured with the hydrogen thermometer, because, as we shall see, the gas would become liquid before this point was reached. We shall use T to represent temperatures measured from absolute zero on the hydrogen scale, called absolute temperatures on the hydrogen scale. On this scale the freezing point of water is 273.04°T, whereas on the absolute thermodynamic scale, also called the *Kelvin scale*, it is 273.19°K. The Kelvin scale is not identical with the hydrogen scale, but the difference (§352) is small. The table on page 198 lists some temperatures of general interest.

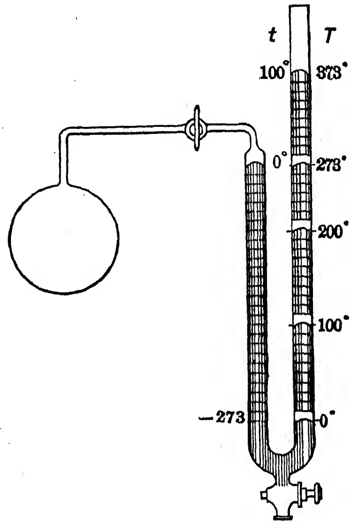


FIG. 171.—Temperature scale determined by change in pressure of a gas at constant volume, P_0 = 1 Atm. = external pressure.

264. Constant Volume Gas Thermometer.—In order to use the constant volume gas thermometer in the simplest way to measure temperatures according to the standard hydrogen scale, the volume of the bulb should be absolutely constant, and all the gas used (including that in the capillary and over the mercury) should be heated to the temperature to be measured. This is impracticable, and hence corrections of the observed readings must be made. Disregarding all corrections, we shall now derive an approximate expression for the temperature of the bulb corresponding to a given pressure reading.

Let P_0 = pressure of hydrogen at the ice-point,
P_{100} = pressure of hydrogen at the steam-point,
t = some other temperature of the bulb, the value of which is to be determined, measured from Centigrade zero.
P_t = pressure of hydrogen at this temperature t.

Then, in accordance with the definition of the degree (§263), we define any temperature t on the Centigrade scale of the constant volume hydrogen thermometer by the following formula:

$$t = \frac{P_t - P_o}{\frac{P_{100} - P_o}{100}} = \frac{P_t - P_o}{bP_o}$$

In this b stands for the fraction $\frac{P_{100} - P_0}{100P_0}$, which, for hydrogen, is

the same as $\dfrac{\text{increase in pressure for 1°C.}}{\text{pressure at 0°C.}}$

This gives at once

$$P_t - P_o = bP_ot$$

or

$$P_t = P_o(1 + bt)$$

an equation which is approximately obeyed by all gases. If t_o is the temperature for which $P_t = 0$, *i.e.*, absolute zero, it is evident that $t_o = -\frac{1}{b}$. Then absolute temperature is given by

$$T = t + \frac{1}{b} = t + T_0,$$

T_0 being the number on the absolute scale corresponding to 0° on the Centigrade scale.

The constant b is called the "coefficient of increase of pressure," or simply the *pressure coefficient*, of a gas; for hydrogen its value is $\frac{1}{273.04}$, hence the value of the absolute zero of temperature on the centigrade constant volume hydrogen scale, as defined above, would be $-273.04°$. The value of b for air and nitrogen also is not very different from $\frac{1}{273}$, so that these two gases would give constant volume temperature scales approximately agreeing with the standard. Nevertheless the *exact* definition of the standard scale as here given is entirely dependent upon the properties of hydrogen. It has been found impossible, however, to use hydrogen above about 1100°C. because of the ease with which it passes through the walls of the metal bulbs which are used at higher temperatures; and under these conditions nitrogen is usually substituted.

265. Constant Pressure Gas Thermometer.—The constant pressure gas thermometer, which makes use of the increase in volume, with increasing temperature, of a gas confined at constant pressure, is convenient for demonstration purposes, though seldom used for precise measurements. As shown in Fig. 172,

the *constant pressure* used is that of the external atmosphere, and the change in volume is proportional to the displacement of a globule of mercury (or other liquid) along a tube of uniform bore.

The *coefficient of expansion*, that is, the ratio $\frac{V_{100} - V_0}{100V_0}$ where V_0, V_{100} are the volumes at 0° and 100°C. respectively (pressure constant), is approximately $\frac{1}{273}$ for hydrogen, air, oxygen and nitrogen, so that an extremely sensitive indicator may easily be obtained.

With a bulb about 10 cm. in diameter and a tube 5 mm. in diameter the displacement of the globule would be about 10 cm. per degree change in temperature of the bulb. The expansion of air when heated is one of the earliest known effects of heat, and the first thermometer, invented by Galileo in 1593, was based on this principle.

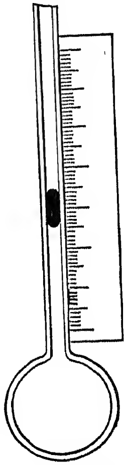


FIG. 172. Constant pressure gas thermometer.

266. Mercury Thermometers.—For ordinary purposes thermometers depending on the expansion of mercury confined in a bulb and tube of glass or other transparent substance are most convenient and universally used. Two standard forms are shown in Fig. 173, the mercury being in a thin-walled glass bulb attached to an extremely fine capillary tube. For use at ordinary temperatures the upper part of the capillary contains only mercury vapor. Since mercury expands somewhat less than $\frac{1}{5000}$ part of its volume at 0°C. for a degree rise in temperature (compare with air above), it is necessary to have a very fine capillary in order to obtain an easily observable motion of the column for a degree change in temperature. All such thermometers should, for precise work, be calibrated or standardized by comparison with the hydrogen standard.

Fig. 173 shows the two standard ways of marking the "scale" on the thermometer. In one the scale is marked directly on the stem of the thermometer. This is the most accurate and permanent way and is used in all standard scientific thermometers and clinical thermometers. In the other the scale is on paper or white glass and is enclosed in an outer glass tube behind the capillary stem. This usually gives more legible scales, but they are somewhat likely to become loose and shift with respect to the capillary. A third method is used for cheap "household" thermometers: in this the thermometer is simply mounted on a support which carries the scale.

The glass used for the thermometer (especially the *bulb*) is of the greatest importance, and in recent years great improvements

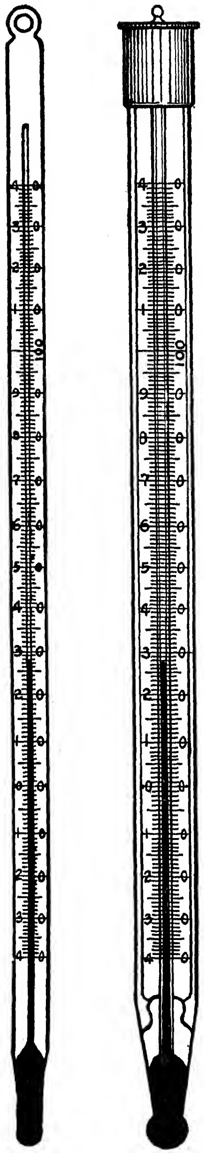


FIG. 173.—Thermometers with solid stem and with enclosed scale.

have been made in the qualities of glass used for this purpose. A bulb of ordinary glass has the fault of slowly changing in volume with time, and of permanently and quickly increasing in volume whenever it is heated, say to 100°C. or higher. Such changes, of course, alter the reading for a given temperature. Some of these effects gradually disappear after the bulb has been made, so that thermometer bulbs should be kept for some time, or else artificially "aged" by heating and cooling, before being graduated.

Through the development of special glasses with high melting-points, it has become possible to construct mercury-in-glass thermometers reading to 550°C. or even higher. In such high-range thermometers the space above the mercury column must be filled with a gas (usually carbon dioxide or nitrogen) at a final pressure of about 19 atmospheres, in order to keep the mercury from boiling. For such thermometers the properties of the glass are of the greatest importance, and the glass known as "Jena 59^{III}" is the best one to use. Even with this glass, if the thermometer is kept at 550°C. for an hour or more, a permanent expansion of the bulb will result. This will permanently lower the freezing-point reading, but if this change is applied as a correction (added) to subsequent readings of the thermometer, fairly correct results can be obtained. Thermometers of mercury in clear fused quartz have also recently been satisfactorily constructed for use up to about 700°C.

In using thermometers it is well to avoid too sudden heating or cooling; and in measurements above 100° (or in all cases where extreme accuracy is required), it must be remembered that thermometers are usually graduated to read correctly when *bulb and stem* are all at the temperature to be measured. If the stem is cooler than the bulb, the thermometer will read too low, and this error may amount

to as much as 40° at 550°C. In careful work thermometers should always be compared with a standard, or standardized at known temperatures (§271) or sent to the Bureau of Standards for comparison.

267. Special Forms of Thermometers.—Alcohol and some other liquids have greater coefficients of expansion than mercury and smaller surface tensions (giving more regular rise and fall in the capillary), but they are seldom used for accurate thermometers. Since mercury freezes at −38.9°C., thermometers containing alcohol are often used for lower temperatures. Pentane (C_5H_{12}) is also used for thermometers reading to −190°C.

Maximum and *minimum* thermometers are thermometers provided with devices for recording the maximum or minimum point

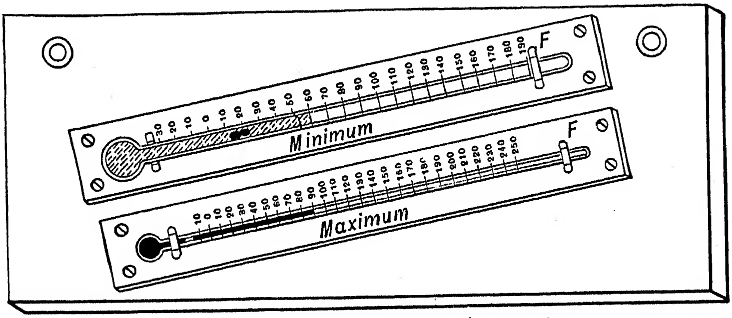


FIG. 174.—Maximum and minimum thermometers.

reached by the end of the mercury column. The maximum thermometer is usually in one of two forms. In the first form a small iron index is pushed ahead of the mercury column and left when the column contracts, the *lower* end of the index indicating the highest reading of the mercury column. In the second form, Fig. 174, there is a constriction in the bore of the tube near the bulb, and at this point the mercury column *breaks*, when contraction occurs after the maximum point is reached, leaving the upper end of the column at the maximum reading. This device is used in clinical thermometers, Fig. 175. Minimum thermometers, Fig. 174, are usually of alcohol in glass, and have below the meniscus a light index, of such a form that the alcohol can flow past it, while it will be dragged *down* when the descending meniscus reaches it. If the thermometer is kept nearly horizontal, the index will rest at the *lowest* point reached by the meniscus.

For some purposes (especially common thermostats) metallic thermometers are used. They usually depend upon the bending of a duplex metallic bar, Fig. 181a, because of the different amounts of expansion of its component metals. They are not satisfactory for accurate work.

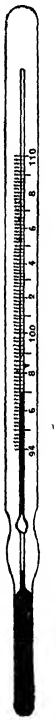


Fig 175.—Clinical thermometer

268. Resistance Thermometry.—In recent years an electrical method of thermometry has come into very general use. In this the thermometric property is the resistance offered by a metallic wire to the passage of an electric current, the resistance changing with the temperature. Such thermometers, like all secondary instruments, must be *calibrated in terms of the standard adopted.* On account of its permanence, high melting-point and acid-resisting qualities, pure platinum wire has been chiefly used for this purpose, though for ordinary temperatures copper and iron wire may be substituted. The usual form of platinum-resistance thermometer is shown in Fig. 176, the coil whose resistance changes are to be measured (called by analogy the "bulb" of the thermometer), being mounted in a protecting tube of glass, or (which is better) of metal for moderate temperatures and of porcelain for high temperatures.

The advantages of the platinum thermometer are permanence and reliability, wide range (it may be used up to 1200°C.), the fact that the readings may be made at a distance of several hundred feet from the thermometer itself and that it may be made accurately self-recording. It is also capable of extreme sensitiveness, $\frac{1}{10000}$°C. being readable. For these reasons its use in scientific and engineering work is rapidly increasing.

Fig. 177 shows the electrical leads to the coil $L_1 L_2$, compensating leads $L_3 L_4$ by means of which the effect of temperature changes in the leads $L_1 L_2$ are eliminated, and the connection of the Wheatstone bridge (see §453) by which the resistance is measured. From an empirical formula developed by Callendar the temperature corresponding to a given resistance may easily be obtained. This formula is of such a form that only three known temperatures are needed to determine its constants. It is, therefore, very easy to standardize a platinum thermometer.

269. Thermoelectric Thermometer.—When two different metals are joined together in a circuit as shown in Fig. 178, and one *junction* is heated, an electromotive force is in general produced (see §464), which tends to drive a current in a certain direction as shown, and this electromotive force increases as the difference in temperature between the two junctions increases. This thermal electromotive force is another thermometric property that is very extensively used. For some purposes a voltmeter (see §438) suffices to measure the electromotive force generated by heating one junction, and it may be calibrated to read temperature directly. The thermoelectric thermometer or thermo-couple, as it is called, is valuable on account of its sensitivity, quick response to temperature changes, and the small size and mass of the part which must be heated as compared to the bulb of a mercury or resistance thermometer.

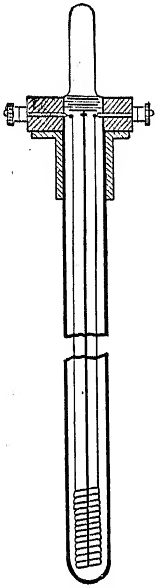


FIG. 176.—Platinum resistance thermometer.

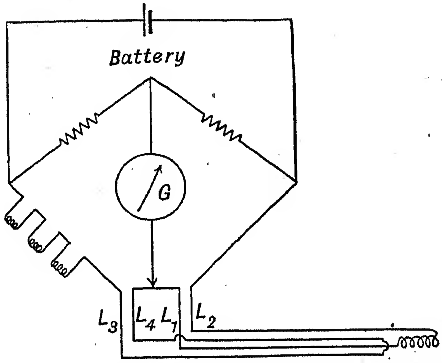


FIG. 177.—Wheatstone's bridge for measuring resistance of platinum thermometer.

For work below 500°C., wires of copper and constantan (an alloy of copper and nickel) are quite satisfactory; up to 1000°C. wires of nickel and nickel-chromium alloy may be used for approximate work; while, for the entire range up to 1600°C., the most accurate results are given by wires of platinum and platinum + 10% rhodium.

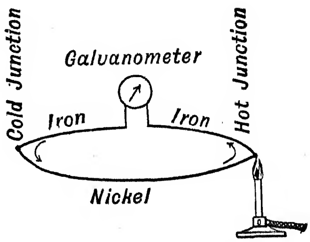


FIG. 178.—Thermoelectric couple, showing direction of current produced by heating.

270. Measurement of High and Low Temperatures.—The measurement of extremely high or low temperatures presents separate and difficult problems. This is partly because of mechanical difficulties caused by changes in properties of ordinary substances at extreme temperatures, for example, melting and softening of metals and porcelain and chemical reactions at high temperatures, and partly because the range of the direct hydrogen thermometer is passed so

that it is necessary to *extrapolate* by means of some empirical formula. At high temperatures recourse is had to nitrogen in a constant volume thermometer, which has been used from 1100°C. to 1550°C.; above this for a short range thermoelectric extrapolation is possible, while beyond this a radiation scale (see §338) and radiation methods are the only resources. At low temperatures the least liquefiable gas, *helium*, used in a constant volume thermometer, but at a pressure of less than 10 cm. of mercury, has been used, as well as the resistance and thermoelectric methods.

271. Standard Temperatures.—For the purpose of standardizing thermometers, thermo-couples, resistance thermometers, etc., it is convenient to make use of one or more temperatures which can easily be obtained and kept constant, and which have been accurately measured. For such purposes melting- and boiling-points are the most convenient. To use a standard boiling-point the liquid must be steadily boiled at a known pressure and the thermometer immersed in the *vapor;* to use a melting-point the thermometer may be immersed in a mixture of the solid and liquid. The following table gives some of the more useful points.

TABLE 2

STANDARD TEMPERATURES

(Pressure Constant at One Atmosphere)

Helium (liquid)	Boiling-point,	−268.9°C.
Hydrogen (liquid)	Boiling-point,	−253°
Oxygen	Boiling-point,	−183
Carbon dioxide	Boiling-point,	− 78.5°
Mercury	Melting-point,	− 38.9
Water	Melting-point,	0
Ether	Boiling-point,	34.6
Alcohol (ethyl)	Boiling-point,	78.3
Water	Boiling-point,	100
Naphthalene	Boiling-point,	218.0
Tin	Melting-point,	231.9
Benzophenone	Boiling-point,	306.0
Sulphur	Boiling-point,	444.7
Sodium chloride	Melting-point,	801
Silver	Melting-point,	960
Gold	Melting-point,	1063
Palladium	Melting-point,	1555
Platinum	Melting-point,	1764
Tungsten	Melting-point,	3393

272. The Pressure, Volume, Temperature Diagram.—From the discussion of §262 we saw that in order to know the *condition* of a body we should know the amount of energy present, per unit mass, in several different forms, namely as kinetic energy of molecules, atoms and electrons and as potential energy of molecules, atoms and electrons. Of the entire amount of this internal energy we have no knowledge, but we can measure the heat energy which passes into or out of a substance and also the external work done, which together constitute the *change* in the internal energy, and hence we can tell when a body is brought

back to a given condition of total internal energy. Now it is found that, in the majority of cases, when a body is brought back to the same total energy content, its pressure, volume and temperature return to the same values, and, in fact, all its physical properties are the same as before; hence it is said that the pressure, volume, and temperature *determine the physical state of a body.*

These three variables, P, V, and t are, however, not independent but are connected by a relation, called an *equation of state,* the general form of which is not known. This relation expresses the experimental fact that, if we fix any two of the three variables, P, V, and t, the third must have a definite value. For example, if a gas occupies a given volume at a given pressure it must have a certain temperature.

Since the physical condition of a body is determined by the values of the three variables, P, V, and t, it is very natural to represent a given condition by a point having the corresponding values of P, V, and t as coördinates measured along three rectangular axes, as in Fig. 179, where every point in space represents a *definite physical condition.* If we take as the origin *absolute zero* values of P, V, and t, then negative values of V and t will mean nothing physically, while negative values of P will mean tensions. Points in a plane parallel to the PV plane will correspond to physical conditions for all of which the temperature is constant, and, similarly, planes parallel to the Vt and Pt planes respectively will represent constant pressure and constant volume conditions. Since it is usually sufficient to fix two of the variables P, V, and t, physical conditions are often represented by points in a plane, for which purpose the PV, Pt, or Vt plane may be chosen.

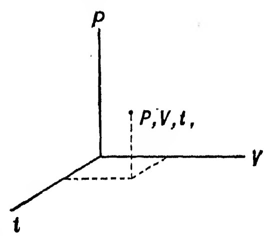


Fig. 179.—The use of P, V, t, as coördinates.

EXPANSION

273. Introduction.—The important changes in substances produced by heat are changes in size, changes in the arrangement of molecules with respect to one another, and changes in state from solid to liquid and gaseous. The difference between solids, liquids and gases has been discussed in §157. Solids in general offer great resistance to change of shape, and their molecules tend to assume a definite arrangement in groups, called crystalline structure, not only in obviously crystalline minerals such as quartz, but in all solids. The existence of such structure is sometimes taken as a test for the solid state, though liquids also can have crystalline properties, and it is difficult to draw a sharp distinction between the two. From the heat standpoint the important matters are that the average molecule in a solid moves about much less than in a liquid or gas, and that the potential energy of the molecules with respect to each other is greatest in the gaseous state; furthermore

the potential energy of a solid, liquid or gas changes with its change of size, or expansion due to heat. In discussing the expansion of solids it is convenient to consider both their change in linear dimensions and their change in volume, while for fluids the latter alone has a meaning.

274. Linear Expansion of Solids.—This is an effect very easily observed and very widely made use of. Telegraph wires which sag in summer are taut in winter; the tires of wagon and locomotive wheels and jackets of large cannon are made too small to slip in place and are put on while expanded by heat, so that when cool and shrunk they have a firm grip. Different solids expand differently for the same change in temperature. A simple experimental arrangement for measuring the amount of expansion is shown in Fig. 180, where *A*, a bar of the material being studied,

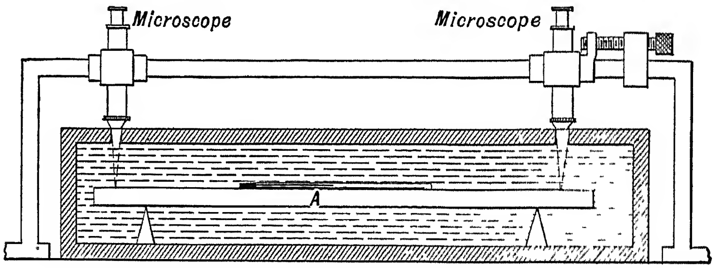


FIG 180—Apparatus for measuring coefficient of linear expansion of solids

is supported in a bath, so that its temperature may be varied. Two microscopes, supported by a frame distinct from the bath, are arranged so that one or both may be moved parallel to the bar by a fine micrometer screw and focused on two fine marks made on the bar. As the bar expands the microscopes are moved so that the cross-hairs remain set on the marks, and thus the expansion can be read from the graduated heads of the micrometer screws. By substituting a standard meter for the bar, the actual length between the marks at any desired temperature, say 0°C., may be determined, and, by adding to this the observed expansions, the length L_t of the bar at any temperature may be obtained. The expansion will usually be found to be approximately, though not exactly, proportional to the change in temperature, that is to say, if the values L_t are plotted as ordinates with the corresponding values of t as abscissæ, the result will be a curve, though the curvature is usually slight. In general it is found that L_t may be very

closely represented by an expression of this form,

$$L_t = L_o(1 + at + bt^2 + ct^3 + \ldots) \qquad (1)$$

where a, b, c are constants and t is the temperature on the Centigrade scale. The number of constants necessary increases with the temperature range over which it is attempted to work and

TABLE 3

Coefficients of Linear Expansion

Substance.	Per degree C.
Aluminum	23.0×10^{-6}
Brass	18.9 "
Copper	16.7 "
Glass (Jena 16^{III})	7.8 "
Glass, Pyrex	3.2 "
Gold	14.2 "
Hard rubber	80 "
Ice	50.7 "
Invar	0.7 "
Iron (cast)	10.2 "
Iron (wrought)	11.2 "
Lead	29.1 "
Nickel	12.8 "
Oak, ‖ grain	4.9 "
Oak, ⊥ grain	54.4 "
Platinum	8.9 "
Porcelain (Berlin)	2.8 "
Quartz, ‖ axis	8.0 "
Quartz, ⊥ axis	13.4 "
Quartz, fused	0.30 "
Silver	18.8 "
Steel	10.5 "
Tin	22.5 "
Zinc	30 "

with the accuracy desired, and it varies also with different substances. For small temperature differences a, usually called "the coefficient of expansion," is sufficient, and its value is evidently

$$a = \frac{L_t - L_o}{L_o t}$$

Frequently also a *mean coefficient of expansion* between two temperatures, t_1 and t_2, is used, and its value is accordingly

$$a_m = \frac{L_2 - L_1}{L_1(t_2 - t_1)}$$

For moderate ranges of temperature (*e.g.*, 0° to 100°) a and a_m usually differ so little that they need not be distinguished.

As may be seen from Table 3 on page 209, the coefficients of expansion are never large, and very refined experimental methods are necessary to determine them accurately, as, for instance, some form of *interferometer* (§699).

Isotropic solids, including crystals in the cubical system (with three equal axes of symmetry), expand equally in all directions. Other crystals have one axis of symmetry, with one coefficient of expansion along this axis and another one in a plane at right angles to the axis, the coefficient being the same in all directions in this plane; while still others have three different coefficients of expansion along three axes, in some cases even showing a contraction along one axis. In such cases of unequal expansions the angles of a crystal change as the crystal expands.

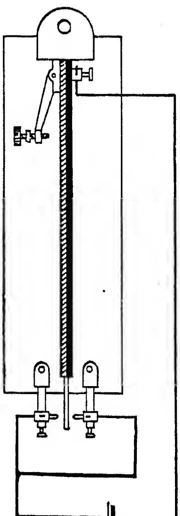


FIG. 181*a*.—Metallic thermometer depending upon difference in expansion of two metal strips, arranged as a thermostat.

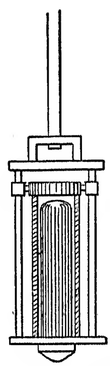


FIG. 181*b*.—Mercury compensation pendulum.

275. Applications of Linear Expansion.—The expansion of solids, especially the differential expansion, is made use of in metallic thermometers, thermographs and thermostats. Usually a compound strip of brass and iron, riveted together, is fixed at one end and arranged so that the bending of the strip, due to the unequal expansion of brass and iron, operates a recording or indicating pointer, or, in the thermostat, makes electrical contact to right or

left and thus controls some heating system. Fig. 181a shows a common form.

The balance wheel of watches has a rim made of a compound metal strip, as above described, and so arranged that a change in temperature, by altering the curvature of these strips, will move a part of the mass of the wheel toward or away from the center, thus altering the rotational inertia of the wheel and hence its period. In this way other temperature effects on the rate of the watch, such as change in elasticity of the springs, change in diameter of the balance wheel, and change in viscosity of the oil in the bearings, may be compensated.

In the mercury clock pendulum shown in Fig. 181b the length of the reservoir of mercury is so chosen that the expansion of the mercury, which raises the center of gravity, just compensates for the expansion of the supporting rod, which lowers the center of gravity, so that the time of vibration will not be altered by changes in temperature. This compensation can now be accomplished even more accurately by the use of a specially worked nickel-steel alloy, called "invar," which has a coefficient of only .00000075 to .00000015, or $\frac{1}{100}$ that of brass. This alloy is also valuable for making standard meter-bars, tapes and scales with lengths practically independent of small temperature variations.

The cracking of objects by heating, particularly sudden heating, is due to unequal expansion produced by differences of temperature in different parts. Porcelain is less liable to crack than glass because of its smaller coefficient of expansion, and thin glass than thick because of the more rapid equalization of temperature. In fusing metals into glass to make an air-tight joint, it is necessary to use a metal having nearly the same coefficient of expansion as glass, otherwise cracking (or leaking) would occur when the joint cooled. As may be seen from the table, platinum is the best metal for this purpose although inexpensive and satisfactory alloys have been developed. There is a very striking difference between the coefficients of crystalline and fused quartz; the former cracks with the slightest heating, the latter, because of its small coefficient, may be taken from an oxyhydrogen flame and at once plunged into liquid air without cracking.

276. Cubical Expansion of Solids.—If V_t represents the volume of a solid at t°C., V_o its volume at 0°C., it is found that in general solids expand in such a way that V_t may be represented as a func-

tion of t by an equation similar to the one used for linear expansion:

$$V_t = V_o(1 + a't + b't^2 + c't^3) \tag{1}$$

and, as before, the constants, b', c', etc., are much smaller than a', so that for small temperature changes,

$$V_t = V_o(1 + a't), \tag{2}$$

If we now consider a cube of the material of length L_t on an edge, we have, approximately, using equation (1), §274,

$$V_t = L_t^3 = L_o^3(1 + at)^3,$$

or,

$$V_t = L_o^3(1 + 3at).$$

Since $V_o = L_o^3$, by comparison with equation (2), we get

$$3a = a'.$$

Hence **the coefficient of cubical expansion is three times the linear coefficient,** and can be obtained from Table 3, p. 209.

277. Expansion of Liquids.—The change of volume of liquids with temperature has already been mentioned as the basis of liquid-in-glass thermometers. The fact that the mercury or alcohol in such thermometers rises with increased temperature shows that the liquid expands more than the glass, and this is usually true of liquids as compared with solids. To represent the volume V_t of a liquid at a temperature t in terms of the volume V_o at 0°C., it is found that an equation of the same form will suffice—

$$V_t = V_o(1 + a''t + b''t^2 + \ldots),$$

or approximately,

$$V_t = V_o(1 + a''t)$$

since b'' is usually much smaller than a''.

A bulb with a capillary stem like a thermometer is usually used in measuring the *differential expansion* of a liquid and a solid. The walls of the bulb expand as if they were filled with solid material; hence the volume of the bulb space is at any temperature equal to the expanded volume of the solid which would fill it at 0°C. If

V_o = volume of bulb or of liquid filling bulb at 0°C.,
V'_t = volume of bulb at t°C.,
V_t = volume of same liquid at t°C.,
a', a'' = volume coefficients of expansion of solid composing the bulb, and of the liquid respectively, then

$$V_t - V'_t = V_o[(1 + a''t) - (1 + a't)] = V_o(a'' - a')t.$$

This *differential or apparent expansion*, $V_t - V'_t$, can be measured by noting the rise of the liquid in the capillary stem. If, in addition, the volume coefficient a' of the solid is known, we can determine the coefficient a'' of the liquid; for

$$a'' = \frac{V_t - V'_t}{V_o t} + a'$$

The difference $a'' - a'$ is called the *apparent coefficient of expansion* of the liquid.

It is possible to determine the *absolute coefficient of expansion* of a liquid, independent of the expansion of a containing vessel, by a method due to Dulong and Petit and illustrated in its simplest form in Fig. 182. Two vertical tubes filled with the liquid in question are connected at their lower extremities by an accurately horizontal tube. The vertical tubes are in baths of some sort, so that one can be maintained at a temperature of 0°C., and the other at t°C. At the bases of the two tubes the pressures must be equal, otherwise there would be a flow from one to the other through the connecting tube. The pressure at the two upper free surfaces must be the same, since it is that of the external atmosphere; hence the difference in pressure from top to bottom of the two columns must be the same. Hence by §185,

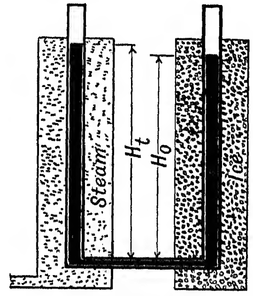


FIG. 182.—Method of measuring the absolute coefficient of expansion of mercury.

$$h_t \rho_t g = h_o \rho_o g$$

and

$$\frac{\rho_o}{\rho_t} = \frac{h_t}{h_o}$$

But if $\quad V_o$ = volume of unit mass of fluid at 0°C. $= \dfrac{1}{\rho_o}$

and $\quad V_t$ = volume of unit mass of fluid at $t°C. = \frac{1}{\rho_t}$

then $\quad V_t = V_o(1 + a''t)$

and $\quad \frac{\rho_o}{\rho_t} = \frac{V_t}{V_o} = 1 + a''t = \frac{h_t}{h_o}$

Hence $\quad a'' = \frac{h_t - h_o}{h_o t}$

TABLE 4

COEFFICIENTS OF CUBICAL EXPANSION OF LIQUIDS

Substance.	Per degree C.	
Alcohol (ethyl)	110	$\times 10^{-5}$
Alcohol (methyl)	118	"
Benzine	124	"
Mercury	18.2	"
Paraffin oil	90	"
Pentane	159	"
Toluene	109	"
Water, 15–100°	37.2	"
Xylol	101	"

This method has been especially used to determine the absolute coefficient of expansion of mercury; this being known, mercury can be used to determine the coefficient of expansion of solids by the differential methods. The coefficients of expansion of liquids (except water) decrease with increase of the pressure at which they are observed.

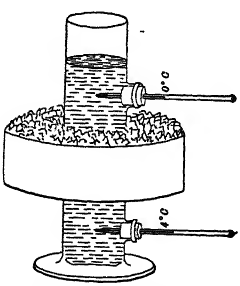


FIG 183.—Hope's apparatus for determining the temperature of maximum density of water.

278. Expansion of Water.—Water is unique among liquids in that it has a maximum density at about 4°C., under 1 atmosphere pressure, *i.e.* below 4°C. it contracts with rise of temperature, above 4°C. it expands.

This property, which has very important consequences, is clearly shown by Hope's apparatus, Fig. 183. If the tank around the middle of the glass vessel be filled with a freezing mixture of ice and salt, and the vessel be filled with water at a temperature higher than 4°C., the water in the middle when cooled will become denser and fall to the bottom, whereas the water above the middle will not be disturbed. Thus the upper thermometer will indicate a practically stationary temperature and the lower one a falling

temperature, until the lower half of the vessel is filled with water at 4°C., after which the upper one begins to fall in temperature until 0° is reached, when freezing may begin at the top, or more probably near the center where the heat is lost to the freezing mixture. The water at 4°C. is most dense and therefore collects at the bottom of the vessel. A somewhat similar operation goes on in winter in ponds and rivers which are not too much disturbed by winds or currents, the densest water, at 4°C., collects at the bottom, while the coldest, at 0°C., being lighter, stays on top. Freezing usually occurs at the top of water as naturally exposed, owing to the loss of heat occurring through the upper surface by exposure to cold air. In some cases, if the entire mass of water is

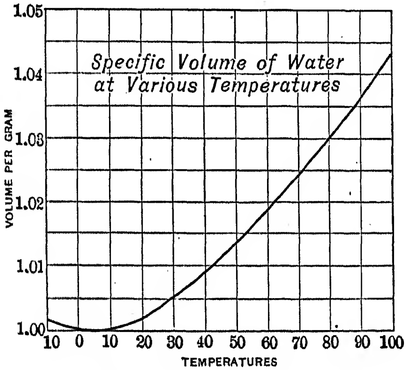


FIG. 184.—Expansion curve for water.

cooled by currents to near 0°C., freezing may occur on the bottom or on submerged solids, cooled by radiation, thus forming *"ground ice,"* which is of serious consequence in northern rivers. The volume of 1 gram of water at various temperatures under 1 atmosphere pressure is given in Fig. 184.

According to Amagat the temperature of maximum density falls with increase of pressure, being about 2°C. under a pressure of 93 atmospheres. If kept by pressure in the liquid state water continues to contract below 0°C. and to expand at an increasing rate above 100°C. The solution of various salts in water also lowers the temperature of maximum density, 4 per cent. of dissolved common salt lowering it to −5.63°C. The peculiar behavior of water as regards its thermal expansion is due, according to Tammann, to the existence at low temperatures of several different kinds of water

molecules or groups of molecules, which gradually break up into one simpler kind as the temperature is raised

279. Expansion of Gases.—Since the effect of pressure on the volume of a gas is very great (§221), it is evident that, in discussing the expansion of gases with increase in temperature, we must be careful to specify the pressure conditions which are to hold during the expansion. The simplest condition is to maintain the *pressure constant* and measure the change in volume of the gas in a bulb by allowing it to expand and push out a mercury piston in an attached tube, as illustrated in Fig. 172. For accurate work an arrangement such as is shown in Fig. 185 is necessary, and for more complete knowledge of the subject the expansion must be carried out at various constant pressures. A correction must, of course, be made for the expansion of the bulb and for the fact that an increasing amount of the gas will be in the stem and hence will not be heated. Gay-Lussac (1802) and Charles (1787) independently carried out such experiments, and arrived at the "Law of Charles and Gay-Lussac," according to which *all the common gases expand by a constant fraction of their volume at* 0°C. *for each rise of* 1° *in temperature.* This fraction is about .003660 ($\frac{1}{273}$), or about the same as the "pressure coefficient" of a gas (§264) In the form of an equation this law is—

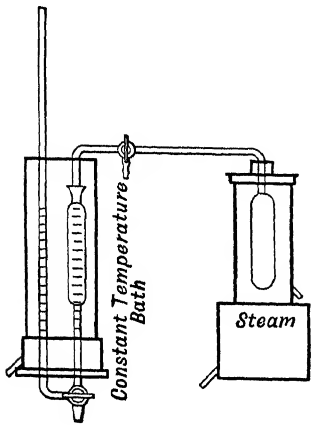


FIG 185.—Apparatus for measuring the expansion of gases The gas expanding from the bulb in steam is measured in the graduated bulb.

$$V_t = V_o(1 + at) \qquad (p \text{ constant}),$$

but it is now known that the law is only approximately true and that a is not the same for all gases. Furthermore, a varies with the pressure and with the temperature, and is not, in general, quite equal to the pressure coefficient, b.

Later work of Regnault and others has shown that the coefficients of expansion of all gases except hydrogen increase, at ordinary temperatures, with increasing density of the gas, and that the coefficients for the several gases are more nearly alike and more

nearly equal to their "pressure coefficients" when the gases are at low pressures or high temperatures.

TABLE 5

EXPANSION COEFFICIENTS AND PRESSURE COEFFICIENTS, PER DEG. C.

Gas.	a.	b.
Air	0.003671	0.003674
Carbon dioxide	0.003722	0.003711
Hydrogen	0.003661	0.003663
Nitrogen	0.003670	0.003671

Temperature 0°C. to 100°C., Pressure 1 atmosphere.

280. The Gas Equation: A Perfect Gas.—We have seen (§221) that gases follow Boyle's law more or less closely, the product of the pressure and volume at constant temperature being nearly constant. In §264 we considered the change in pressure with

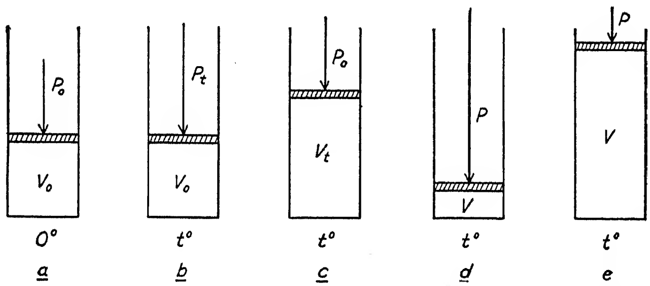


FIG. 186.—Pressure and volume relation. The product PV is the same for cases b, c, d, e.

temperature of a gas confined at constant volume, which is given approximately by the equation $P_t = P_o(1 + bt)$. In §279 we have just discussed the expansion of gases, which follows approximately the relation $V_t = V_o(1 + at)$, the pressure being constant. It is convenient to combine these statements into a single equation, which will then represent all the relations which approximately hold between the pressure, volume and temperature of a gas. This may be done as follows:

Let P_o and V_o (Fig. 186) be the pressure and volume of a given mass of gas at 0°C., and let it be heated at constant volume V_o to t°C. Then we have (§264)

$$P_t = P_o(1 + bt) \qquad (1)$$

and hence

$$P_tV_o = P_oV_o(1 + bt)$$

Again, starting at P_o, V_o, $0°$, let it be heated at constant pressure, P_o to the same final temperature $t°$C ; then by the law of Charles (§279)

$$V_t = V_o(1 + at) \tag{2}$$

and hence

$$P_oV_t = P_oV_o(1 + at)$$

But the final temperature of the gas is $t°$C. after each of these operations. Hence by Boyle's Law the product of the pressure and the volume must be the same for the final conditions and equal to the product of the pressure and volume for any condition at this temperature, that is

$$P_oV_t = P_tV_o = (PV)_t \tag{3}$$

or

$$P_oV_o(1 + at) = P_oV_o(1 + bt)$$

from which

$$a = b \tag{4}$$

and also

$$(PV)_t = P_oV_o(1 + bt)$$

But from §264

$$t + \frac{1}{b} = T$$

so that

$$(PV)_T = P_oV_obT$$

or in general

$$PV = RT \text{ (where } R \text{ stands for } P_oV_ob) \tag{5}$$

This equation may also be put in the form $PV/T = R$. Then, if a given mass of gas is subjected to different conditions as regards pressure, volume and temperature,

$$\frac{P_1V_1}{T_1} = \frac{P_2V_2}{T_2} \tag{6}$$

This equation is very useful for computing one of these quantities when the others are given.

It is frequently convenient to consider an imaginary *ideal* or *perfect* gas, which *exactly* obeys these gas laws, and which also has certain other properties that will be referred to later. The volume and pressure coefficients of such a gas we shall designate by a' and b', absolute temperature according to this *perfect gas* scale by T', and the constant factor by R'. The gas law or equation of state for a perfect gas then becomes $PV = R'T'$.

281. Real Gases.—As we have just seen, real gases follow more or less closely the law

$$PV = RT$$

where T is the temperature measured with a constant volume hydrogen thermometer from the absolute zero of the hydrogen scale. The approximation to the law $PV = RT$ is found to be very much closer for high temperatures and low pressures.

The value of R is different for different gases, and, of course, also for different masses of any one gas. It is customary to consider the equation as applying to 1 gram of gas, and R is then called the *gas constant* for this gas. It is evidently equal to $\frac{1}{273.2}$ of the product P_oV_o at 0°C. For any other mass of M grams the constant in the equation $PV = RT$ will be MR, since volumes are proportional to masses under given conditions.

It is easy to see, in a general way, why the properties of real gases should approach those of a perfect gas at high temperatures and low pressures. For, according to the simple kinetic theory (§227), a gas having no molecular forces, *i.e.*, *no molecular potential energy*, and *negligible molecular volume*, is perfect in so far that it obeys the law $PV = RT$. Now, it is evident that the higher the temperature of a real gas, the less will be the proportion of the potential to the kinetic energy, and also that the larger the volume of a gas, other things being equal, the less will be the actual molecular volume compared to the total volume. Hence, as the temperature is raised, or the density diminished, the conditions become more nearly those assumed in the simple kinetic theory.

By making a still further assumption equation (5) of §280 may be further generalized. According to Avogadro's law equal volumes of different gases at the same temperature and pressure contain

equal numbers of molecules, that is, the total masses of equal volumes will be proportional to the molecular weights of the gases, or

$$M_1:M_2:M_3 \quad .\ . = m_1:m_2:m_3 \ . \ . \ .$$

where m_1, m_2, m_3, are molecular weights. Hence, if we take m_1 grams, m_2 grams, and m_3 grams (called gram-molecular-weights) of these gases, they will occupy the same volume at the same pressure and temperature.

Hence, $\qquad PV = m_1R_1T = m_2R_2T = m_3R_3T$
and, $\qquad m_1R_1 = m_2R_2 = m_3R_3 = R''$

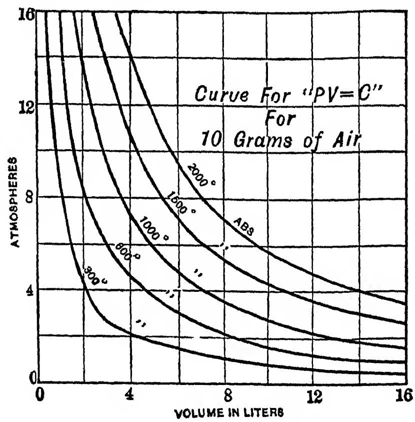


FIG 187 —Isothermal curves for air.

where R'' is a constant for all gases, the value of which can be at once computed. For example, for nitrogen $m = 28$; specific volume $V = 796.2$ c.c. when $T = 273°$: and $P = 1$ atm. $= 1{,}1013250$ dynes/cm.^{2}
Hence

$$R'' = \frac{PVm}{T} = 8.305 \times 10^7 \frac{\text{ergs}}{\text{degree}}$$

282. Isothermal Curves.—The significance of the equation $PV = RT$ can be seen more readily by graphical representation according to the method of §272 When T has some constant value, T_1, Boyle's law, $PV =$ const., is represented by a rectangular hyperbola in a plane parallel to the PV plane and cutting the T axis in the point T_1. If a series of such hyperbolæ are located

for different temperatures and then *projected* upon the PV plane by dropping perpendiculars from every point to this plane, the result is a *family* of hyperbolæ, each of which can be distinguished by labeling it with the temperature belonging to it, as shown for air in Fig. 187. Any curve showing the relation between the *pressure* and *volume* of a substance under the condition T = const. is called an *isothermal* curve. We accordingly conclude that the isothermal curves for a perfect gas are rectangular hyperbolæ, and that isothermal curves for real gases approximate to rectangular hyperbolæ, the approximation being closer at high temperatures.

283. Molecular Energy and Temperature.—We have seen (§221) that ordinary gases very approximately obey Boyle's law, PV = constant for constant temperatures, and that PV increases as the temperature increases. Also, according to the kinetic theory of gases (§227), for a simple ideal gas $PV = \frac{1}{3}M\overline{v^2}$ which will be constant if the average random undirected kinetic energy per molecule is constant, and will increase in proportion to the average molecular energy $\frac{1}{2}M\overline{v^2}$. From these two statements for a real and an ideal gas it is natural to conclude that the temperature of a real gas is, at least approximately, proportional to the kinetic energy of molecular motion, and even to extend this analogy to liquids and solids where it cannot be justified so simply. While the proportionality of mean molecular kinetic energy to temperature turns out to be very closely true for gases, and the relation is very useful and instructive, the complicated structure of real molecules (as compared with those of the ideal gas) indicates that the relation is probably not, in all cases, so simple.

CALORIMETRY

284. Unit of Heat.—Calorimetry is the process of measuring quantities of heat. Obviously the first thing to be decided upon is the unit in terms of which to measure. Though, as has been said, energy units may be used, it is often more convenient to use a unit defined in terms of *heat phenomena* only.

In looking for a purely thermal unit of heat it is natural to pick out some effect which heat produces, and agree that the heat unit shall be such an amount of heat as will produce a *specified* amount of this effect in unit mass of a *standard substance*. The *specified effect* agreed upon is a change in temperature of 1°C., and the *standard substance* is water. To be exact the particular

degree must be specified; hence we shall define the unit of heat as *that quantity of heat which will raise the temperature of* 1 *gram of water from* 14.5 *to* 15.5°*C.* This is called the calorie, or cal_{15}.

The relation of this thermal unit to the unit of mechanical energy has been found by experiments which will be described later (§339). These show that if the "mechanical equivalent of heat," that is, the number of work units equivalent to one heat unit, be denoted by the letter J,

$$J = 4.185 \times 10^7 \frac{\text{ergs}}{\text{calorie}}.$$

Sometimes a *mean calorie* is also specified. This is one one-hundredth of the heat required to change 1 gram of water from 0° to 100°C. It is about equal to 1 cal. Sometimes the "large calorie," equal to 1000 calories, is used as a unit. In engineering practice (in English-speaking countries) the "British thermal unit" (B.T.U.) is employed; it is equal to the heat required to raise the temperature of 1 lb. of water 1° Fahrenheit. From the relation of the pound to the gram and the Fahrenheit to the Centigrade degree, it follows that:

$$1 \text{ B.T.U.} = 252 \text{ cal.}$$

In British thermal units and foot pounds J is 778 ft-lbs./B.T.U.

The most common method of measuring quantities of heat in calories is the "method of mixtures," which consists in transferring the quantity of heat to be measured to a known mass of water and observing the resulting rise of temperature of the water. The heat may be transferred to the water in many ways—for example, by dropping a piece of hot copper into the water, by pouring some hot liquid into it, or by passing steam into it. It is, of course, simplest to use the water at about the temperature for which the calorie is defined, as in that case the number of grams of water used multiplied by the number of degrees rise in temperature will give at once, to a first approximation, the number of calories which have been added.

285. Specific Heat.—If two different masses of water are exposed for the same length of time in just the same way to a steady source of heat, it will be found that the temperatures of the two have risen inversely in proportion to their masses. If the same masses of copper be treated in the same way, the rise in temperature

will be more than ten times as great, but again inversely proportional to the masses. From this we conclude that the temperature effect of a given heat agent, acting on a body for a given time, depends on the mass of the body and on a factor which differs for different substances, and which is called the *specific heat*. The *specific heat of a substance is defined as the heat required to raise the temperature of* 1 *gram of it* 1°C. under constant pressure. The symbol for specific heat is s. To be exact, the particular degree must be specified, because the specific heat varies with the temperature, for example, the number of calories required to raise 1 gram of a substance from 0° to 1° is different from the number required to raise it from 49° to 50°. For most purposes, however, and for not too large temperature differences, say from 0° to 100°, it is not necessary to consider

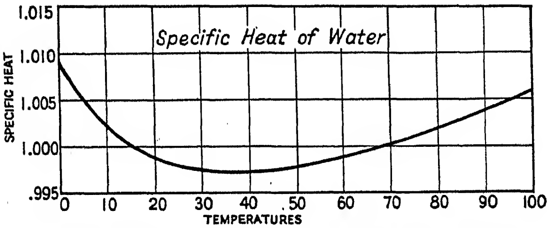


FIG. 188.—Variation of specific heat of water with temperature.

the *variation* in specific heat, and it is customary to speak of *the* specific heat, meaning the mean value within the range considered.

The *heat capacity*, S, of a *body*, of any mass and variety of parts, is the number of calories required to raise its temperature 1°C. at the mean temperature t. This will evidently depend on the masses and specific heats of the various parts of the body, and if m_1, m_2, m_3 and s_1, s_2, s_3, stand for the masses and corresponding specific heats of the parts, we have

$$S = m_1s_1 + m_2s_2 + m_3s_3 + \ldots \text{ etc.}$$

286. The Variation of the Specific Heat of Water.—The common occurrence of water and its physical and chemical characteristics make it extremely useful in heat measurements; hence a knowledge of its specific heat at various temperatures is of importance. The specific heat of water is, of course, unity at the temperature for which the calorie is defined (§284). At other temperatures it may be either greater or less than unity. The first satisfactory study of the variation of the specific heat was made by Rowland

in 1878; combined with later work, it shows that the specific heat diminishes with rising temperature, reaching a minimum about 35°C. The mean value of the specific heat of water from 0° to 100°C. differs very little from 1.

287. Method of Mixtures.—Returning now to a more detailed consideration of the method of mixtures, we find that there are in practice several additional points to be considered, as can best be seen by discussing a particular form of apparatus shown in Fig. 189. In the first place the water must be held in some vessel C, containing a stirrer and a thermometer and called a *calorimeter*. Into this some of the heat will pass, raising its temperature. Moreover, some heat will pass out of the water and containing vessel during the operation and will, therefore, fail to produce its proportionate temperature change. To take account of the first effect we must know the heat capacity of the calorimeter. The second effect, loss of heat to the surroundings, necessitates what is called the *cooling* or *radiation correction*. Neglecting this correction for the moment, we can write the fundamental equation of the mixture calorimeter thus:

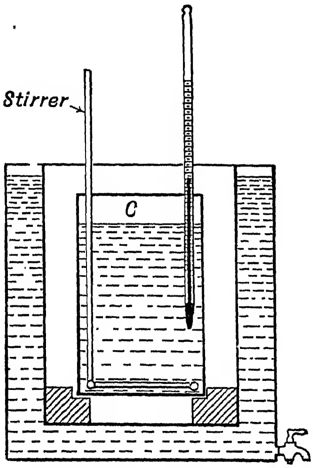


FIG. 189.—Calorimeter for method of mixtures.

$$H = (ms' + \Sigma m_1 s_1)(t_2 - t_1)$$

which expresses the fact that the heat added, H, equals the heat gained by the calorimeter and the water. In this expression s', s_1, s_2, etc. are respectively, the mean specific heats of water and of the various materials of the calorimeter, m is the mass of the water, $\Sigma m_1 s_1$ the heat capacity of the calorimeter and stirrer, etc., t_2 the final and t_1 the initial temperature of the calorimeter and water.

Unless special precautions are taken, the loss of heat to the surroundings, which is largely due to convection (§323) rather than radiation, is relatively great, and the cooling correction is a very important one. In general it can be reduced by protecting the calorimeter from air currents, polishing its exposed surface, surrounding it with a constant temperature enclosure, and arranging matters

so that t_1 and t_2 are respectively slightly below and slightly above the temperature of the enclosure. In some cases it is impossible or inconvenient to use water as the calorimetric substance, and then some other liquid or solid of known specific heat may be used.

TABLE 6

SPECIFIC HEATS

(Calories per Degree C. per Gram)

Substance	Specific heats	Temperature C.
Alcohol (ethyl)	0.548	0°
Aluminum	0.219	15 to 185
Aluminum	0.0093	−240
Brass	0.090	0
Copper	0.0936	20 to 100
Copper	0.00036	−250
Diamond	0.113	11
Diamond	0.0003	−220
Glass (flint)	0.117	10 to 50
Gold	0.0316	0 to 100
Granite	0.19 to 0.20	0 to 100
Graphite	0.160	11
Graphite (acheson)	0.0573	−79 to −190
Ice	0.502	−21 to −1
Iron	0.119	20 to 100
Lead	0.0305	20 to 100
Lead	0.0143	−250
Mercury	0.0333	20
Mercury (solid)	0.00329	−40 to −75
Nickel	0.109	18 to 100
Platinum	0.0323	0 to 100
Quartz	0.166	0
Silver	0.0559	0 to 100
Sodium	0.2433	−83 to −190
Tin	0.052	19 to 99
Turpentine	0.420	18
Sea water	0.980	17
Zinc	0.0935	0 to 100
Zinc	0.0017	−240

288. Application of Method of Mixtures.—The method of mixtures may be used for different purposes, according to the source of the heat H which is to be measured. One important use is in determining the specific heat of substances. For this purpose a known mass M of the substance is heated to a tempera-

ture t (above or below t_1) and placed in the calorimeter and water. The temperature of the calorimeter and of the mass M will then equalize, and, if we call t_2 the final temperature of the mixture, the heat H added to the calorimeter is the heat lost by the mass M in changing from the temperature t to t_2, which, from the definition of specific heat, is equal to $Ms(t - t_2)$, where s is the *mean* specific heat of the substance M in the interval t to t_2. We then have:

$$H = Ms(t - t_2) = (ms' + \Sigma m_1 s_1)(t_2 - t_1)$$

from which s can be computed.

Sometimes it is advisable to keep the hot body from direct contact with the water by putting it in an inner vessel having thin walls of good conducting material.

The method of mixture is also used to determine heats of fusion (§305) and evaporation (§312), as well as the amount of heat developed or absorbed in various chemical reactions. In such cases the operation consists in fusing, or condensing, or combining, as the case may be, known masses of material inside the calorimeter (in the inner vessel above referred to). Special forms of calorimeters, called combustion calorimeters, bomb calorimeters, etc., have been developed for these purposes.

289. Method of Continuous Flow.—A second method for measuring quantities of heat is the method of "continuous flow," illustrated in Fig. 190. A steady stream of the calorimetric substance (usually water from a reservoir) at a constant temperature is allowed to flow past the point at which heat is being set free, in such a manner that all of the heat is absorbed by the stream of water. The temperature of the stream of water is, of course, higher after the heat has been absorbed than before, and if the rate of liberation of heat is constant, this temperature difference will be constant, and (neglecting external losses as before) the number of calories liberated in a time T, *since it does nothing but heat the water,* will be equal to the number of grams of water M which has flowed past in time T, multiplied by the number of degrees rise in temperature $(t_2 - t_1)$, and by the specific heat of water s', or

$$H = Ms'(t_2 - t_1)$$

This method is especially useful in determining the heats of combustion of gas and liquid fuels, by means of which a steady rate of combustion and hence a steady liberation of heat can be maintained. This method can also in a sense be *reversed* by generating the heat mechanically or electrically, that is, by converting measured amounts of mechanical or electrical energy completely

into heat, which is absorbed by a stream of fluid whose specific heat is to be determined. The equation then becomes,

$$H = Ms(t_2 - t_1)$$

where H is known (from mechanical or electrical measurements) in energy units, M is the mass of fluid flowing past in time T and s is its specific heat, which is determined by this equation in mechanical units. In this form the method has been used by Barnes to measure the specific heat of water and mercury, and it is capable of giving very accurate results.

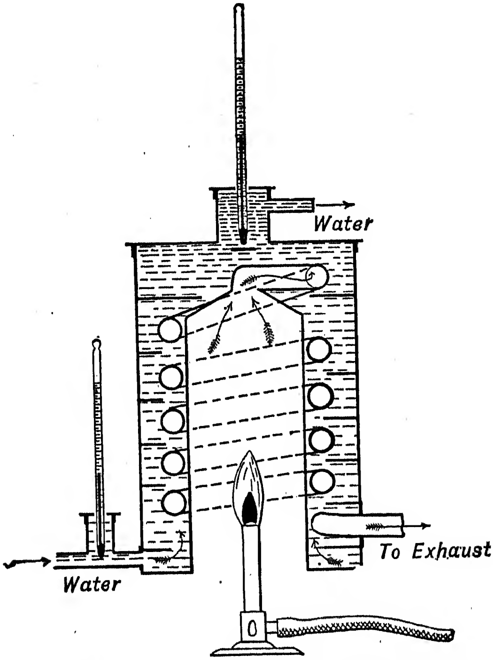


Fig. 190.—Continuous flow calorimeter for measuring heat of combustion of gas.

290. A third method of measuring a quantity of heat is the "method of latent heats," in one form of which the heat to be measured is used to melt a measurable amount of ice. This necessitates a knowledge of the amount of heat required to melt 1 gram of ice (heat of fusion of ice, §305), but has the advantage that the calorimeter remains at a fixed temperature, 0°C. The most common instrument of this type is the Bunsen ice calorimeter. A second form is the Joly steam calorimeter, in which the order of temperatures is reversed, and the amount of heat required to raise M grams of a substance from a temperature t to the temperature of steam, say 100°C., is determined from the weighed amount of steam which is condensed to supply this heat. A knowledge of the heat liberated in condensing 1 gram of steam is, of course, necessary. This is a very convenient and reliable method.

291. The Specific Heat of Gases.—In the previous discussion of specific heat we have neglected one factor which, as we have seen in §279, becomes very important as soon as we consider gases, namely, the expansion which usually accompanies rise in temperature. If a gas is confined in a cylinder with a movable piston, as, for instance, in a bulb with a mercury plug in an attached capillary tube (Fig. 172) and is heated, it will, as we have already noted, expand and push out the mercury plug. The outside of the plug is acted upon by the pressure of the air, which opposes its motion outward by a force equal to the product of the pressure and the cross-section of the tube. Overcoming this force through a given distance means doing work, called the *external work of expansion,* and this work has evidently been done by the expanding gas.

Looking at the matter from the standpoint of the kinetic theory, we should say that, before the confined gas was heated, the impact of gas molecules on the inner end of the mercury plug (at rest) was balanced by the impact of air molecules on the outer end, but that an increase in temperature of the gas meant more and harder impacts on the inner end, thus destroying the equilibrium and causing the plug to move. The *moving* plug would, on the average, hit the outside molecules harder than it had previously done when at rest; hence it would increase the velocity of these molecules and add kinetic energy to them. On the other hand, the plug would be *moving away* from the inside molecules, so when they hit it they would rebound with slightly *less* energy. The work done by the expansion consists therefore in a transfer of kinetic energy from the gas molecules inside to air molecules outside.

Thus, if heat energy imparted to the confined gas causes expansion, some energy will be, by this expansion, taken out of the gas, and this is a possible disposition of part of the added energy quite separate from those considered in §262. Hence we can see that to raise the temperature of a gas with the volume kept constant must take an amount of energy different from that required if expansion against pressure is allowed, not only because in the second case the internal potential energy may be increased but also because external work is done. In other words, the heat added to a gas (or any body) is equal to the increase in *internal* kinetic and potential energy plus the *external* work done.

The amount of both the internal and the external work will evidently depend on the amount of expansion. Since the increase

in volume of gases per degree rise in temperature is very much greater than that of solids or liquids, the external work is also greater. If the volume is not kept constant it may be allowed to vary in many ways, the most important being such an increase

TABLE 7

SPECIFIC HEATS OF GASES AND VAPORS

Substance	Temperature	Specific heats		s_P/s_V
		s_P	s_V	
Alcohol (ethyl)	108–220	.453	.400	1.133
Air	20–440	.237	.....	1.402
Argon	20– 90	.123	.....	1.667
Benzine	34–115	.299	.214	1.397
Carbon dioxide	15–100	.2025	.....	1.299
Chlorine	16–343	.113	.....	1.336
Chloroform	27–118	.144	.125	1.152
Ethyl ether $(C_2H_5)_2O$	25–111	.428	.....	1.024
Helium	18	1.251	.....	1.63
Hydrogen	12–198	3.409	2.42	1.408
Mercury vapor	310	....	.....	1.66
Nitrogen	20–440	.242	.171	1.41
Oxygen	20–440	.224	.....	1.398
Water vapor	100	.480	.....	1.33

in volume that the *pressure remains constant.* Hence we have the specific heat of a gas at constant volume, s_V, and the specific heat at constant pressure, s_P, defined as the heat necessary to raise the temperature of 1 gram of the gas 1°C. under the condition of constant volume or constant pressure respectively. From what has been said it is evident that s_P must be greater, in general considerably greater, than s_V.

The measurement of s_V has been most accurately made by means of the steam calorimeter (§290), a known mass of the gas being enclosed in a metallic bulb, and the weight of steam condensed in raising it from t°C. to 100°C. being determined; a correction must then be made for the thermal capacity of the bulb. s_P is usually measured by passing a stream of heated gas through a calorimeter. According to Regnault and later observers, s_P for most gases varies only slightly with pressure, while s_P for air is almost independent of the temperature, but for CO_2 increases very

markedly with temperature. s_V for air and CO_2 increases with the density of the gas. The value of s_V has not been determined directly for many gases, but the value s_P/s_V can be readily deduced from the velocity of sound in the gas (§561).

292. The Free Expansion of a Gas.—We have already seen that, if there are forces between molecules and atoms, when a gas expands there will be a change in the potential energy of its molecules (and perhaps of its atoms), since the average distance between molecules will increase. Work done against internal forces in this way is called the *internal work* of expansion to distinguish it from the *external work* done against the pressure confining the gas.

Gay-Lussac and, later, Joule attempted to measure the internal work by the method of free expansion (Fig. 191) in which gas was confined at some considerable pressure in the vessel A and allowed to expand quickly through a cock C into B, which had been highly exhausted. A, B, and C were in a vessel of water whose temperature was measured. Since expansion occurred into a vacuum it was "free" (no opposing pressure) and hence, on the whole, no external work was done; but if any internal work had been done, it should have caused a change in the temperature of the gas. For if the internal potential energy increases, the kinetic energy must decrease by an equal amount, that is, the temperature of the gas must fall, and *vice versa* (§283). Joule did not measure the temperature of the gas but that of the water, and its large heat capacity so masked the effect that his results merely indicated that *the internal work of expansion of a gas is small.*

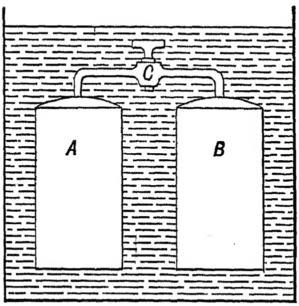


FIG. 191.—Illustrating Joule's study of the "free expansion" of gases.

293. Temperature of Gas in Motion.—If a gas at high pressure and ordinary temperature is allowed to escape into the atmosphere through a fine tube (Fig. 192) in which it acquires a high velocity of flow, very marked cooling effects will be observed where the velocity of flow is greatest, though the total energy of the moving gas is practically the same as that of the gas at rest. The explanation is that part of the energy of the random undirected motion of the particles, which determines the temperature, has become temporarily energy of directed motion in the stream.

But if the gas is caught in a large receiver and allowed to come to rest, its temperature will be found to be slightly higher than before expansion. If A, B, and C of Fig. 191 are placed in separate vessels, it will be found that the expansion lowers the temperature of A and raises that of B by almost an equal amount. This heating and cooling is partly due to the fact that relatively more of the fast-moving molecules will escape, and partly to the fact that the gas moving out of A corresponds, with respect to the gas remaining in A, to the piston moving *away from* the gas in Fig. 172. Hence the gas remaining does work on the gas which is set in motion, and the one loses and the other gains heat of equal amounts.

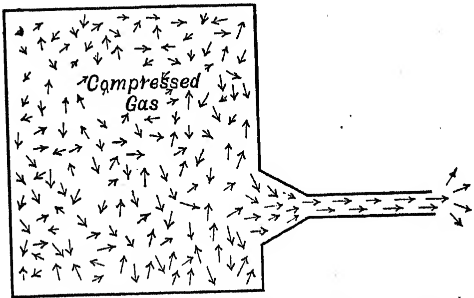


FIG. 192.—The change from undirected to directed molecular motion in an escaping gas.

294. The Difference between the Two Specific Heats.—From the definition of s_P and s_V in §291 and from the statements made in §292 we see that if we denote the external work of expansion by W_e and the internal work of expansion by W_i per deg. per gm.

$$s_P = s_V + W_e + W_i$$

The expansion is that necessary to maintain P constant while t rises 1°C., and the two quantities of work must be expressed in heat units. For gases which are not approaching liquefaction, for example for O_2, H_2, N_2, and air, at ordinary temperatures, the internal work of expansion is so small that it may usually be neglected. Then if P is the constant pressure, and ΔV the change in volume per degree per unit mass, the external work is $P\Delta V$ ergs (§195). Dividing by the mechanical equivalent of heat J (§284), this external work is $P\Delta V/J$ calories,

and
$$s_P - s_V = \frac{P\Delta V}{J}$$

Also,
$$PV = RT \qquad \text{approximately,}$$

and $$P(V + \Delta V) = R(T + 1)$$
Hence $$P\Delta V = R$$
and $$s_P - s_V = \frac{R}{J}$$

This equation was used by Robert Mayer in 1842 to make the first computation for J, the other quantities being determined by experiment.

295. The Ratio of the Two Specific Heats.—From the preceding, and taking into account the internal work of expansion, the ratio of the two specific heats is evidently;

$$\frac{s_P}{s_V} = \frac{s_V + \frac{W_i}{J} + \frac{R}{J}}{s_V}$$

Also, according to the kinetic theory, s_V = increase in molecular kinetic energy + increase in atomic energy, per degree: the first we shall denote by E_m, the second, which is the increase in energy *inside* the molecule, we shall represent by E_a. Then from §283

$$PV = \frac{M\overline{v^2}}{3} = \frac{2}{3} \cdot \frac{1}{2}M\overline{v^2} = RT.$$

Hence $$E_m = \frac{1}{T} \cdot \frac{1}{2}M\overline{v^2} = \frac{3}{2}R$$

so that $$s_V = \left(\frac{3}{2}R + E_a\right)\frac{1}{J}$$

and $$\frac{s_P}{s_V} = \frac{\frac{5}{2}R + E_a + W_i}{\frac{3}{2}R + E_a}$$

If E_a and W_i are both relatively small, as we should expect them to be in simple monatomic gases which approximately obey Boyle's law, for example argon, then

$$\frac{s_P}{s_V} = \frac{5}{3}, \text{ approximately.}$$

On the other hand if all the other terms are negligible compared with E_a, as we might expect for gases with very complicated molecules, for example ether,

$$\frac{s_P}{s_V} = 1 \text{ approximately.}$$

296. Expansion against Pressure.—Let a gas be forced through a small aperture in such a way that the pressures before and after passing the opening are maintained constant. A possible way of doing this is shown in Fig. 193, in which the pistons both move to the right as the gas passes through, and the external forces upon them are constant.

Let P_1 = pressure of gas before expansion.
V_1 = specific volume before expansion.
P_2 = pressure of gas after expansion.
V_2 = specific volume after expansion.

Then the external work done upon the gas by the first piston while unit mass is passing is P_1V_1 (§195), the external work done by the gas after expansion upon the second piston is P_2V_2, and $(P_2V_2 - P_1V_1)$ is the net amount of external work, W_e, done by the gas, and this may be either positive or negative. The apparatus is supposed to be so made that no heat can enter or leave the gas during the operation, and the temperature of the gas is observed before and after expansion.

Let W_i again represent the internal work done against the forces between molecules, and Δt the observed change in temperature of the gas. Then the sum of the external and internal work must be equal to the change in the kinetic energy of the molecules plus any change in the energy inside the molecules, and this, as was just shown in §295, is s_V. If both W_e and W_i are positive, that is if some of the kinetic energy of the molecules goes into both external and internal work, the temperature must fall, hence we have the following equation.

$$-\Delta t s_V = W_e + W_i$$

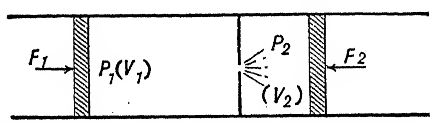


FIG. 193.—The Joule-Kelvin porous plug experiment. Unbalanced but not "free" expansion.

If the gas strictly obeyed Boyle's law and if there were no temperature change, P_1V_1 would be equal to P_2V_2, that is W_e would be zero, and hence, by the above equation, the internal work would be zero. As a matter of fact, with O_2, N_2, and CO_2 a cooling is observed, with H_2 at ordinary temperatures a heating, and from these observations, combined with the value of the specific heat s_V and the variation of PV, the internal work, which Joule could not detect, can be computed. The results indicate in all cases *molecular forces of attraction*. To avoid false cooling effects due to mass motion of the gas (§293), it is customary, following the plan of Lord Kelvin, to use many fine openings—that is a *porous plug*—hence the experiment is known as the "porous plug experiment."

297. Relations between Specific Heats.—In view of the complicated nature of molecular structure as already outlined, it is evident that no simple relations are to be expected between the specific heats of different substances or of the same substance at various temperatures, and the following general statements must suffice.

The best known attempt to express a relation between the specific heats of substances is the so-called "law" of Dulong and Petit (1819), which states that the "product of the specific heat by the atomic weight is the same for all solid elements." But this statement and its extension to compound substances are only rough approximations.

In general the specific heat of substances in the liquid state is much greater than in the solid (two times as great for water, ten times for mercury), while s_p for the gaseous state is about the same as that of the solid. The change of specific heat from solid to liquid is in most cases smaller with metals than with non-metals.

With most substances, solid, liquid or gaseous, the specific heat increases with rise of temperature, though the change is small for solids, with the exception of carbon, boron and silicon. This variation of the specific heat may, according to §262, be due to several causes, such as an increase in the relative amount of kinetic energy inside the molecule and atom, increase in the number of free electrons, or an increase in the potential energy of molecular groups or groups of atoms. The change of molecular grouping is probably the chief cause of variation. The fact of variation with temperature shows at once that Dulong and Petit's "law" cannot be a general one.

Nernst extended the measurement of specific heats down to 23° absolute (−250°C.), and showed that the specific heats of all the substances examined decrease very greatly at extremely low temperatures (Table 6). He also developed new relations between the specific heats of elements and compounds which are of great importance.

TABLE 8

Heats of Combustion. (Calories per Gram.)

Substance	Heat of combustion	Substance	Heat of combustion
Alcohol (ethyl)	7183	Gas (coal gas)	5800–11000
Alcohol (methyl)	5307	Gas (illuminating gas)	5200–5500
Benzine	9977	Gunpowder	730
Carbon (diamond)	7860	Hydrogen	34100
Carbon (graphite)	7800	Petroleum (Am. crude)	11100
Coal (anthracite)	7800	Wood (beech)	4168
Coal (bituminous)	6100–7800	Wood (oak)	3990
Coal (coke)	7000	Wood (pine)	4420

298. Heats of Combustion.—A very important use of calorimeters is in measuring heats of combustion of fuels, that is, the heat liberated by the burning (in air or oxygen) of 1 gram of coal, wood, oil, gas, etc. Such fuels are the source of the larger part of the available energy of the world, and a knowledge of the energy available per unit mass of the fuel is, of course, of great importance to the engineer. The method of mixtures is usually used for solid fuels, especially with one form of apparatus called a "bomb calorimeter," in which a weighed amount of fuel is enclosed with compressed oxygen in a steel bomb and ignited electrically, the

bomb being in the water of the calorimeter. For liquid fuels and gases a method of continuous flow (§289) is also very much used. The heat of combustion is usually expressed in calories per gram, or B.T.U. per pound of fuel.

CHANGE OF STATE

299. Change of State.—The most marked changes in the physical properties of bodies occur when they change from the solid to the liquid or gaseous state.

The change from the solid to the liquid state is called *fusion* or *melting*, the reverse change, *freezing*.

The change from the liquid to the gaseous state is called *vaporization*, the reverse change, *condensation*.

The change from the solid directly to the vapor state is called *sublimation*, the reverse change, *condensation*.

Each of these changes involves a rearrangement of the molecules with respect to each other, and perhaps a rearrangement of the atoms and electrons forming the molecules. Vaporization and sublimation also involve a very great increase in the average distance between molecules. Rearranging and separating the molecules will involve an increase in potential energy in passing from the solid to the liquid and to the gaseous state, while any change in volume will involve doing external work (§292); hence energy must be added to the body to bring about the change. Conversely, when a vapor condenses or a liquid solidifies, a certain amount of energy must be taken away from it. As groups of liquid molecules "settle down" into the solid arrangement, some of their potential energy becomes kinetic and is given up to the surface on which freezing occurs. Thus we have the general relation that:

Energy necessary for fusion or vaporization = increase in potential energy of molecules + external work.

300. Fusion.—If a crystalline solid is heated while acted upon by a constant pressure, it will begin to melt at a definite temperature called the fusing-point or melting-point, and the entire mass will remain at this temperature until it is all melted. To determine this temperature, a thermometer bulb of some kind, protected by a metal or porcelain tube, may be put in the mixture of solid and liquid, as in Fig. 194. Or a thermometer may be placed in a molten substance which is allowed to slowly lose heat; the temperature will fall until solidification begins, after which it will

remain constant, while potential energy (heat of fusion) is being given up, until solidification is complete. The constant temperature is the *freezing-point* of the substance.

The freezing-point of water is found by immersing a thermometer in a mixture of pure ice and water, carefully protected from gain of heat from the outside. As has been stated, the *freezing-point of water under one atmosphere pressure* is one of the fixed points of thermometry.

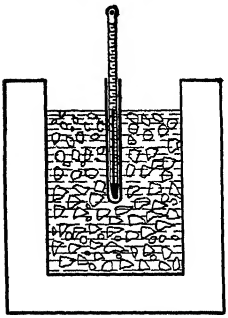


FIG. 194.—Method of determining a freezing-point.

301. Effect of Pressure on Fusion.—If a substance contracts on melting, it is obvious that external work will be done *on* it during the process of fusion, and this will supply some of the energy necessary. Hence an *increase* in external pressure, since it increases the external work, will facilitate melting, so that it will occur at a lower temperature. Conversely, *increase* in pressure *raises* the melting point of substances which expand on melting. This relation between change of melting-point with pressure, and change of volume on fusion, was first deduced from theoretical considerations by James Thomson. The curve obtained by plotting melting-points and the corresponding pressures on the *Pt* diagram is called the fusion-curve (Fig. 200); it represents the conditions under which the solid and liquid can exist in equilibrium in contact. The change in the melting-point with pressure is in general not large, being for water 0.0072°C. *decrease* in temperature for each atmosphere increase in pressure. This decrease does not continue indefinitely. There is a minimum freezing point at −22°C. under 2500 atmospheres pressure. As a result of the work of Bridgman and Tammann it is now known that water can exist in several different solid forms, according to the pressure applied. In one form of ice the melting point has been followed up to 80°C. at 20,000 atmospheres pressure.

The lowering of the freezing-point of water with pressure has important consequences and may be strikingly illustrated. Two blocks of ice at about 0°C. will freeze together if two faces are pressed together; snow compressed in a cylinder becomes a clear transparent mass of ice, and snow at about 0°C. can be pressed by the hands into a hard snow ball. If a wire supporting a weight

is looped around a block of ice, it will slowly melt its way through the ice, which freezes again about it. A further illustration is the fact, well known to skaters, that ice is more slippery when near 0° than when many degrees below zero. In all these cases the pressure applied, which may be very considerable at certain points, lowers the melting-point, and, if the initial temperature of the snow or ice is not too low, some of it will melt, only to freeze again when the pressure is relieved. Thus there would be a film of water between the skate runner and the ice. If the ice or snow is too cold, the pressure will not lower the melting point below this initial temperature and no melting will occur. The same ideas apply to the "packing" of snow on roads, and on a larger scale to the formation of glaciers by the compression, melting, and *regelation* of snow in mountain valleys. The subsequent flow of glaciers down the valleys is due in part to the effect here discussed, the ice melting at the points of greatest pressure, the water immediately flowing down hill a little, thus relieving the pressure and then freezing again.

TABLE 9

TABLE OF MELTING-POINTS

Substance	Melting-point	Substance	Melting-point
Aluminum	658.9°C.	Mercury	−38.9°C
Copper	1083	Nickel	1452
Gold	1063	Platinum	1764
Helium (at 25 atm. pressure	−272.2	Silver	960
Iridium	2350	Tin	232
Iron	1535	Tungsten	3393
Lead	327	Zinc	419

302. Crystalline and Amorphous Solids.—The sharp change from solid to liquid at a definite temperature, which we have been discussing, is characteristic of solids that have a definite crystalline structure. Solids that have not such a structure, called amorphous solids, of which fats, waxes, glass and most alloys are examples, change gradually from one state to another, that is, gradually soften throughout the entire mass, while the temperature rises slightly, there being no definite "melting-point." Amorphous solids are in general mixtures. Some alloys, however, are definite *compounds*, having marked crystalline structure and very definite melting-points. The freezing-point curve for a simple group of alloys is shown in Fig. 195.

303. Change of Volume on Freezing.—Most substances contract on freezing, the solid sinking in the liquid. The fact that iron, bismuth, antimony and some alloys such as type-metal (lead, antimony, and tin) expand on solidifying is valuable industrially, since, when cast, they take a particularly sharp impression of the mold. The expansion of water upon freezing is responsible for the bursting of water pipes, the bursting (and hence death) of plant cells, and the splitting of trees and rocks. Very carefully dried seeds may be put in liquid air without injury, but the presence of the slightest trace of moisture will result in killing the seeds.

304. Freezing Point of Solutions.—A dilute solution, such as sea water, has a lower freezing-point than the pure solvent, and the lowering of the freezing-point of dilute solutions is approximately

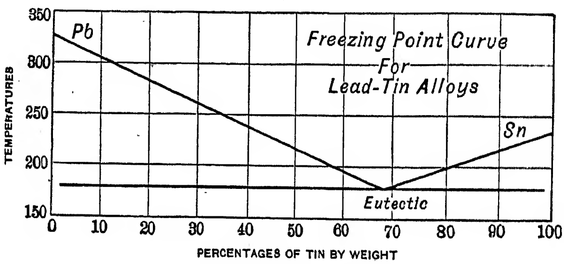


FIG. 195.—Freezing-points of alloys. Upper line, beginning of freezing; lower line end of freezing; eutectic has a sharp freezing-point.

proportional to the amount of substance dissolved. The depression of the freezing-point per gram-molecule of solute in 1000 grams of solvent, calculated from observations on dilute solutions, is called the *molecular lowering of the freezing-point.* The freezing-point of a given solvent is lowered the same amount by many different solutes in proportion to their molecular weights, while other solutes produce a depression two or three times as great.

According to the dissociation hypothesis, abnormally large depressions are due to the breaking up, or dissociation, of molecules into parts, while abnormally small depressions are due to the grouping together of molecules. Thus common salt in water apparently dissociates into Na and Cl, giving a solution which conducts electricity readily and producing a molecular lowering of the freezing-point of about 3.6°. If the temperature of a given dilute solution is lowered beyond the freezing-point corresponding to its saturation, the pure solvent only will begin to freeze out of the solution. Thus

alcohol and water in a radiator first freezes as a sort of mushy mixture of ice and liquid. The remaining solution becomes more concentrated, until, on continued cooling, a certain definite concentration is reached (depending upon the pressure) and the entire mass freezes as a mixture of the two solids. This mixture is called a *cryohydrate*. The corresponding mixture in the case of alloys that have a minimum melting-point as compared with other percentage compositions is called a *eutectic*. (Fig. 195.)

TABLE 10
HEATS OF FUSION
(Calories per gram)

Aluminum	77
Copper	43
Ice	79.6
Lead	5
Mercury	3
Platinum	27
Sulphur	9
Zinc	26

305. Heat of Fusion.—The *heat of fusion* of any substance is defined as **the number of calories required to convert one gram of the solid at the melting-point into liquid at the same temperature.** Heats of fusion are usually measured by some modification of the method of mixtures.

Thus if M = number of grams of melted substance used,
t_3 = temperature of substance when added to calorimeter,
t_m = melting point of substance,
t_2 = final temperature of calorimeter,
t_1 = initial temperature of calorimeter,
m = mass of water used,
$\Sigma m_1 s_1$ = heat capacity of calorimeter,
s_l = specific heat of substance when melted,
s_s = specific heat of substance when solid,
L = heat of fusion,

then

$$(m + \Sigma m_1 s_1)(t_1 - t_2) = M[s_s(t_m - t_3) + L + s_l(t_2 - t_m)]$$

from which L may be computed.

306. Vaporization.—From the molecular standpoint, vaporization means the flying off of molecules against the forces of molecu-

lar attraction, these molecules losing kinetic energy and gaining potential energy as they leave the liquid. The more rapidly moving molecules will be the first to fly off; hence the average kinetic energy of the molecules remaining behind will be less than the initial average for the liquid, and the liquid will be *cooled* by evaporation. If the vapor is confined over the liquid, some vapor molecules will strike the surface and become liquid again, and as the number of vapor molecules per unit volume (*i.e.*, the density of the vapor) increases, the number of molecules returning to the liquid per second will likewise increase, until finally the average number returning will equal the average number leaving. Under these conditions *the vapor is in equilibrium with the liquid.* The density, and hence the pressure, of the vapor when there is equilibrium will depend on the temperature, that is, on the average molecular velocity. A vapor in equilibrium with the liquid is said to be *saturated,* and the equilibrium pressure is called the *saturated vapor pressure* (or *vapor tension*); for a given substance

TABLE 11

VAPOR TENSIONS AND VAPOR DENSITIES OF WATER

Temperature	Vapor tensions (mm. of Hg.)	Densities of saturated vapor (grams of vapor per cu. m. of saturated air)
−20°C.	0.770	0.894
−10	1.947	2.158
0	4.579	4.85
10	9.205	9.41
20	17.51	17.30
30	31.71	30.35
40	55.13	51.10
50	92.30	83.20
60	149.2	130.5
70	233.5	198.4
80	355.1	293.8
90	525.8	424.1
100	760	598
140	2710	1968
180	7514	5150
260	35187	....
360	139880	....

it depends practically only on the temperature. If the vapor is not allowed to accumulate over the liquid, it will remain unsaturated, equilibrium will not be reached, and the liquid will gradually disappear by evaporation.

No general relation is known connecting the saturated vapor pressure and temperature, though many empirical relations have been found and are satisfactory in certain cases. The corresponding values of temperature and saturated vapor pressure for water are shown in Table 11 and Fig. 196. Points in this diagram indicate the physical condition of water substance; points on the curve show the conditions under which water may exist either as a vapor or liquid or both in equilibrium, as, for example,

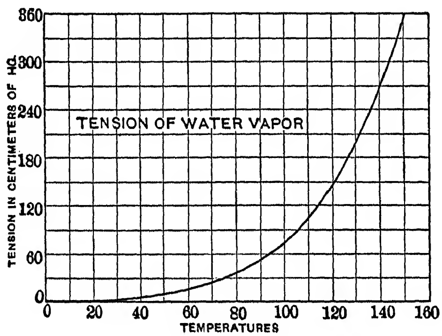


Fig. 196.—Pressure (tension) of saturated water vapor at various temperatures.

at a temperature of 140°C. and under a pressure of 270 cm. of mercury. If the pressure is increased, without suitably raising the temperature so as to reach another point on the curve, all the vapor will be condensed, while if the temperature is increased without properly increasing the pressure, all the water will vaporize. Hence this curve, which represents equilibrium conditions, divides other conditions into two groups, an all-vapor group, represented by points to the right of and below the curve, and an all-liquid group, represented by points to the left of and above the curve.

307. Humidity.—The saturated vapor pressure for a given temperature is not measurably affected by the presence of gases that do not chemically combine with the vapor. When we speak of air being saturated with water vapor, what we really mean is that the vapor is saturated. The presence of air above a water surface will not change the vapor pressure necessary for equilibrium,

but will affect the rate of evaporation if the equilibrium condition is not reached.

The degree of saturation of air with water vapor is of great importance in its influence upon climate, for it determines the rate at which evaporation will go on from exposed surfaces of water or from moist surfaces, such as that of the human body. Evaporation, as we have seen, causes cooling; hence the less saturated the air the greater the cooling, since evaporation will be more rapid. Fanning or blowing upon a moist surface blows away the blanket of saturated vapor which tends to form about it. The rate of evaporation and consequent cooling effect is accordingly increased. A given summer temperature with the air dry is less oppressive than with the air nearly saturated. The effective dryness of air depends on its approach to saturation, and this is called the humidity, the *relative humidity* being defined as the ratio of the mass of moisture actually present in any volume to the amount needed for saturation at the same temperature, while *absolute humidity* is the mass of water vapor contained in a cubic meter of air

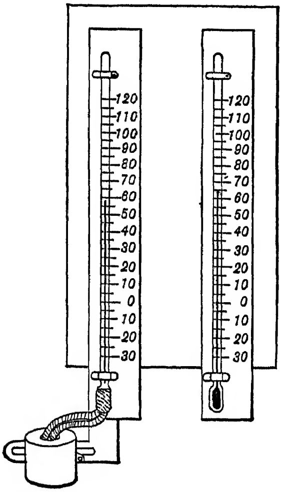


FIG 197 —Wet and dry bulb hygrometer

If water vapor (or air and water vapor) is heated at approximately constant pressure, without the addition of vapor, as in a hot air furnace, it expands, and therefore the mass of vapor per unit volume decreases. Moreover at the higher temperature the density necessary for saturation is greater, so that for two reasons the effective dryness is increased. As might be expected, the air in houses in winter is usually far too dry for either comfort or good health.

The measurement of humidity is called *hygrometry*. The *wet and dry bulb hygrometer* (Fig. 197) consists of two exactly similar thermometers similarly exposed, except that the bulb of one is covered with a light wick kept moist by dipping in a vessel of water. Evaporation from the wet bulb will cool it as compared with the other, and the dryer the air the greater will be the difference in temperature between the two By noting this difference and the temperature of the dry bulb, either the relative or the

absolute humidity may be obtained from tables. Air must circulate freely around these thermometers in order that they should give accurate results.

Much more reliable results are obtained if the wet and dry bulb thermometers are swung rapidly through the air, this form being called the "sling psychrometer." If the instrument, the wet bulb in particular, is not fanned rapidly or swung through the air, the rate of evaporation and cooling effect will be diminished by the blanket of nearly saturated vapor. Other methods depend upon the determination of the *dew-point,* that is, the temperature at which air would be saturated by the moisture actually in it. One form of dew-point hygrometer, Regnault's, is shown in Fig. 198. The central chamber into which the thermometer dips is cooled by the evaporation of the ether contained in it, and the temperature at which condensation first occurs on the front polished metal face is the dew-point. At ordinary room temperatures the proper relative humidity is from 60 to 70 per cent. To reach this humidity, it is necessary to evaporate water in the furnace or in the rooms themselves. If evaporation is desired in the furnace, it is evident, from what has been said above, that the *hot* air (not the *cold* air as is usually the case) should pass over the water surface in order to take up as much water vapor as possible and not be "dried" by subsequent heating. To give 70 per cent. relative humidity at 68°F. in a house 40 ft. by 30 ft. and 25 ft. high, containing about 30,000 cu. ft., requires about 22½ lbs. or 2.8 gallons of water, and usually several times this amount must be supplied per day to maintain proper conditions. The attainable or desirable indoor humidity depends, more or less, on the climate. In cold climates a winter indoor relative humidity of more than about 40 per cent is likely to produce excessive condensation on the windows.

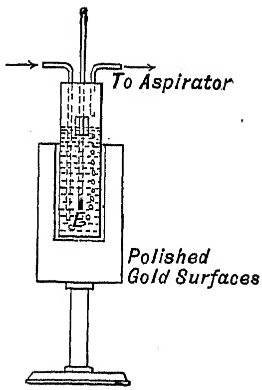


Fig. 198.—Regnault's dew-point hygrometer.

308. Boiling-point.—The boiling-point of a liquid is the temperature at which its saturated vapor pressure is equal to the atmospheric pressure on the liquid surface. At this temperature bubbles of vapor will form in the liquid and escape through the surface, and this formation and escape of vapor bubbles is called *boiling.*

Evidently the boiling-point will be higher the greater the pressu on the surface of the liquid, and the variation of the boiling-poi with pressure is the same thing as the change of saturated vap pressure with temperature. In the determination of boiling-poin the thermometer is put in the vapor rather than in the liquid, ar it must be protected from gain or loss of heat by radiation and fro the condensation of liquid upon it.

TABLE 12

BOILING-POINTS UNDER ONE ATMOSPHERE PRESSURE

Substance.	Boiling-points.
Alcohol (ethyl)	78.3°C.
Benzol	79.9
Carbon dioxide	−78.5
Chloroform	61.2
Ether (ethyl)	34.6
Iron	2450
Mercury	357
Oxygen (liquid)	−182.9
Pentane	36.1

309. Effect of Pressure on Boiling-point.—The variation of the boiling-point with pressure may be determined by enclosing the liquid to be boiled, reducing the air pressure on the surface and noting the temperature at which steady boiling takes place. This may be done by placing the flask under the receiver of an air pump. If the space over the liquid contains only the saturated vapor, and the pressure of this is suddenly reduced, for example by pouring cold water over a sealed flask containing

TABLE 13

CHANGE OF BOILING-POINT OF WATER WITH PRESSURE

Pressure (in mm. Hg.).	Boiling-point	Pressure (in mm. Hg.).	Boiling-point.
680	96.91°C.	740	99.25°C
690	97.32	750	99.63
700	97.71	760	100
710	98.11	770	100.37
720	98.49	780	100.73
730	98.88		

water and its vapor slightly below the boiling point, boiling will at once begin. The pressure variation of the boiling-point of water is frequently used to determine the air pressure on mountain tops and hence roughly their height. The variation for water is about 0.37°C. at 100° for a change in pressure of 1 cm. of mercury.

310. Other Conditions Affecting Boiling.—The ease with which vapor bubbles are formed in a liquid depends upon various conditions, such as the presence of dissolved gases or of points on small solid particles in the liquid. Hence the prevention of "bumping," or the violent formation of vapor bubbles, is brought about by putting broken glass or other rough solids into the boiling liquid. The boiling-point of a given pure liquid is always raised by dissolving any relatively non-volatile substance in it, as, for example, sugar in water, but may be either raised or lowered by dissolving a volatile substance in it, as, for example, alcohol in water, which gives solutions boiling below the boiling-point of water but above that of alcohol. With volatile combinations the boiling-

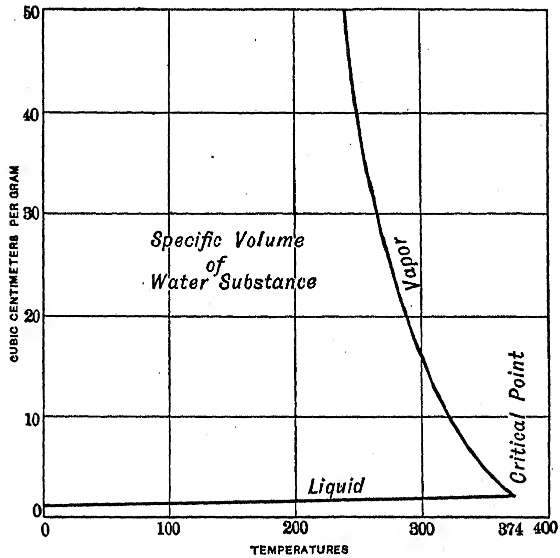


FIG. 199.—Specific volumes of liquid and saturated vapor at various temperatures.

point may be either above or below the boiling-point of *both* constituents, while with a non-volatile solute the elevation of the boiling-point is, for dilute solutions, proportional to the mass of solute added and approximately the same for equal gram-molecular weights of all solutes in the same amount of a given solvent. For water the elevation is at the rate of 0.5° for each gram-molecular weight dissolved in 1000 grams of water, and this is called the *molecular rise of the boiling-point.*

311. Specific Volumes and Densities of Vapors.—In general there is a very large increase in volume on vaporization. Thus 1 gram of saturated steam at 100°C. occupies 1676 c.c., while 1 gram of water at 100°C. occupies only 1.043 c.c. The volume of 1 gram of a substance is termed its specific volume and is evidently equal to $1/\rho$, where ρ is the density. As the temperature

is raised the specific volume of any saturated vapor decreases whereas that of the liquid increases. Fig. 199 shows this behavior for water and steam. The two specific volumes of liquid and vapor become *equal* at a definite temperature, called *the critical temperature*, which will be further discussed in §317.

312. Heat of Vaporization.—The heat of vaporization is the number of calories required to change 1 gram of a substance from a liquid at a certain temperature to a vapor at the same temperature under a specified pressure. The symbol for it is L_t. The method of mixtures is usually used to determine the heat of vaporization, vapor from a boiling liquid being passed into a vessel immersed in the water of a calorimeter, where the vapor is condensed and the heat given off. If M grams of vapor at a temperature t_3 in condensing raise m grams of water and the calorimeter (heat capacity $\Sigma m_1 s_1$) from t_1 to the final temperature t_2 of the mixture,

$$M[L_t + s(t_3 - t_2)] = (m + \Sigma m_1 s_1)(t_2 - t_1),$$

from which L_t can be computed, if s, the specific heat of the condensed vapor, is known.

Since there is in general a very considerable increase in volume on vaporization, which occurs against a definite pressure, *external* work must be done by the vapor as it is formed. The heat of vaporization may, therefore, be considered as made up of two parts, the *internal*, which is the increase in potential energy of molecules and atoms, and the *external*, which is the work done by expansion.

The heat of vaporization diminishes with increasing temperature and becomes zero at the critical temperature, where, as we shall see (§318), the distinction between liquid and vapor vanishes. The heat required to change 1 gram of a substance from a liquid at 0°C. to a vapor at a temperature t is called the *total heat* of the vapor at temperature t. If H_t is the total heat, and we assume the specific heat is independent of temperature,

$$H_t = st + L_t$$

As a liquid evaporates, either the remaining liquid will be cooled or heat will be drawn from the surroundings—hence the cooling effect of evaporation. The rate of evaporation will depend upon the rate of supply of heat; hence boiling over a fire will be more violent in a metal than in a poorly-conducting vessel. It is the heat of vaporization of steam that is chiefly effective in steam heating systems, this heat being supplied in the boiler and given up

(potential converted to kinetic energy again) in the radiators where the steam is condensed. The use of a hand or electric fan in warm weather is the most common example of the reverse process, that is, cooling resulting from increased evaporation due to the circulation of the air.

TABLE 14

HEAT OF VAPORIZATION

(Calories per gram at normal boiling points.)

Alcohol (ethyl)	205	Liquid air	50
Liquid H_2	107	Liquid CO_2	96
Liquid O_2	51	Mercury	70
Liquid N_2	48	Water	540

313. Kinetic Theory Considerations.—The phenomena discussed under Change of State are really very complex because of clustering of molecules and for other reasons, and an adequate explanation would be equally complicated. A simple kinetic theory point of view, however, may help in reaching a general understanding of them.

In the crystal state, which is typical of the solid state, since practically all solids are more or less crystalline, the molecules (or atoms or ions) are regularly arranged in a latticelike structure. Although constrained by forces of attraction to definite positions, the particles are nevertheless vibrating; and when their kinetic energy reaches a sufficiently large value the molecules are torn apart from one another and the structure is broken up, that is, the solid is melted. The energy which must be added to accomplish this break-up is the heat of fusion. Depression of the freezing point, as when ice is in contact with salt, may be explained by the attraction of the salt molecules for the water (ice) molecules, which aids in the break-up of the crystal structure. Accordingly less thermal agitation energy is required, that is, the melting or freezing point is lowered.

As already stated (§310), the boiling point of a pure liquid is raised by dissolving in it a nonvolatile substance, *e.g.*, sugar in water. In this case the attraction between the sugar and water molecules makes the escape of the latter into the vapor more difficult. For a given temperature this results in a lower saturated vapor pressure than would be given by the pure liquid, which means that to give a vapor pressure of one atmosphere the temperature must be higher than normal, *i.e.*, the boiling point is raised.

Kinetic theory gives us the law that for a given temperature the average kinetic energy of different molecules is the same, regardless

of size of molecule or state (solid, liquid, gas) of the substance. This law applies even for particles much larger than molecules, which accordingly have the same energy as the molecules themselves. When minute particles of gamboge mixed with water (or smoke particles in air) are examined under a high power microscope they are seen to dart about in an erratic manner; the smaller the particle the more rapid its motion. This chaotic motion is known as the *Brownian Movement*, after the botanist Brown who discovered it in 1827. Measurements of the energy of such particles allow a determination of n, the number of molecules in a cubic centimeter of gas. For the pressure of a gas is given (§227) by $p = \frac{1}{3}nmV^2$, and since we know the value of mV^2 from measurements of the energy of Brownian particles, we can get n at once. For 1 atmosphere and 0°C. $n = 2.70 \times 10^{19}$

314. Sublimation.—The direct change from the solid to the vapor state is termed *sublimation* This, or the reverse, direct condensation, is commonly observed in the evaporation of snow in cold dry weather, in the production of hoar frost, and in the evaporation and recondensation of camphor confined in a bottle. As in the other two transformations, there is, for every temperature, a definite vapor pressure at which the solid and vapor can exist in equilibrium, without any continuous change from one state into the other. When plotted on the *Pt* plane these points give the "hoar frost" or sublimation line, the equilibrium pressure falling with the temperature. Sublimation also involves an increase in potential energy and external work, and hence a *heat of sublimation.*

315. Unstable Conditions.—In stating that substances solidify, vaporize, and condense at *definite temperatures* under a given pressure, we have disregarded certain cases in which it is possible to cool a liquid very much below its freezing-point without solidification, and to heat a liquid very much above its boiling-point without vaporization. These, however, are abnormal and unstable conditions, since, if freezing (or boiling) once starts, it goes on with great violence till the normal temperature has been restored. For example, small drops of water in an oil of equal density have been cooled at atmospheric pressure to −20°C. without freezing and heated to 178°C without boiling, and minute spheres of platinum and other metals have been cooled several hundred degrees below their normal melting point before solidification occurred. Agitation, the presence of points or particles, or the least trace

of the solid serve to start solidification, while points and pieces of porous solids start boiling. A liquid above its boiling-point will begin to boil violently if touched by a file or paper, though these materials are ineffectual when they are *clean.*

Condensation of a vapor is difficult to start without the presence of *nuclei,* that is, dust particles, liquid droplets, or electrified molecules or atoms called *ions,* which very greatly assist the formation of large drops. The efficacy of some of these is due to their providing surfaces of relatively larger radius on which condensation can take place, since the vapor pressure necessary for the equilibrium of a vapor with a liquid drop is less for a drop of large radius and also less for electrified drops.

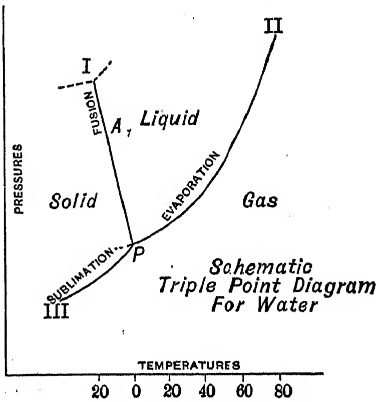


Fig. 200.—Triple point diagram.

316. The Triple Point Diagram Having discussed the three equilibrium curves, solid-liquid, liquid-vapor, and solid-vapor, let us consider them combined as in Fig. 200 in the *Pt* plane. Since the areas on either side of each curve represent conditions such that only one state can exist, for example solid to the left and liquid to the right of the freezing-point curve, a little consideration will show that, in order to be consistent, the three curves must *intersect in a point.* If this were not so, the area included between the curves would represent quite contradictory conditions as deduced from the several curves. Since each curve represents conditions of possible coexistence of two states, in the condition represented by the point of intersection all three, solid, liquid, and vapor can exist simultaneously in equilibrium, and this is the *only* condition in which it is possible. It is called the *triple point,* and for water it corresponds to +0.0072°C. and a pressure of about 4.6 mm. of mercury.

It may be thought that common experience contradicts this conclusion as to the unique properties of the triple point—ice, water, and water vapor being often found coexistent at various temperatures. Under these conditions it will be found, however, that the ice is melting and the vapor condensing, or water evaporating—at least *equilibrium does not exist.* The characteristics represented by the triple-point diagram are presumably true for all substances, though helium behaves in an exceptional way near the triple-point.

317. The Critical Temperature.—In 1822 Cagniard de la Tour made important discoveries regarding the relation between the liquid state and the vapor state by filling a glass tube with alcohol

and its vapor, sealing off the tube, and then heating it. If about two-thirds of the total volume were liquid at ordinary temperatures, he found that, as the temperature increased, the meniscus, or curved surface separating liquid from vapor and due to molecular attractions (§208), became flatter and less distinct and finally disappeared. Thus above a temperature of about 243°C. the liquid and vapor have the same molecular attractions (no meniscus) and are visually identical.

The limiting temperature at which a separation can be observed between the liquid and the vapor state is called the critical temperature. We have already seen that the specific volumes of a liquid and its vapor become more nearly equal as the temperature is raised, and that at the same time the latent heat of vaporization becomes less, and the relation of these facts will now be clear.

The strong attraction that exists between molecules in the liquid state is sufficient to prevent the escape of molecules of average energy, but those of sufficiently high velocity escape into the vapor state. At the critical temperature the molecular kinetic energy is sufficient to overcome the attraction of the molecules in the liquid state and pull them apart from each other. Whatever the pressure may be, the liquid state is impossible; the substance acts like a gas and completely fills the volume.

318. Isothermal Curves for CO_2.—Andrews, in 1863, inclosed CO_2 in a glass tube, kept this at a constant temperature and compressed the gas by forcing mercury into the tube. He measured the pressure required to compress the gas to various measured volumes while the temperature was kept constant, and did this at a series of temperatures from 13° to 48°C. Corresponding pressures and volumes plotted on the PV plane gave what we have called *isothermal curves*, shown in Fig. 201.

The effect of compression at 21.5° from an initial volume of 12 c.c. per gram is, as shown by the curve, first a gradual increase in pressure until a pressure of 59 atmospheres is reached (at A), when liquid CO_2 will suddenly appear in the tube, after which no increase in pressure will occur, in spite of diminution in volume, until B is reached. During this time condensation has continued, until (at B) the vapor has been entirely changed to liquid, after which any decrease in volume necessitates a very great increase in pressure, the liquid being in general very incompressible.

If the same sequence of operations is carried out at a higher temperature, it is found that condensation begins at a less volume and higher pressure, and ends at a greater volume, that is, the *volume interval* during which the substance is part vapor and part

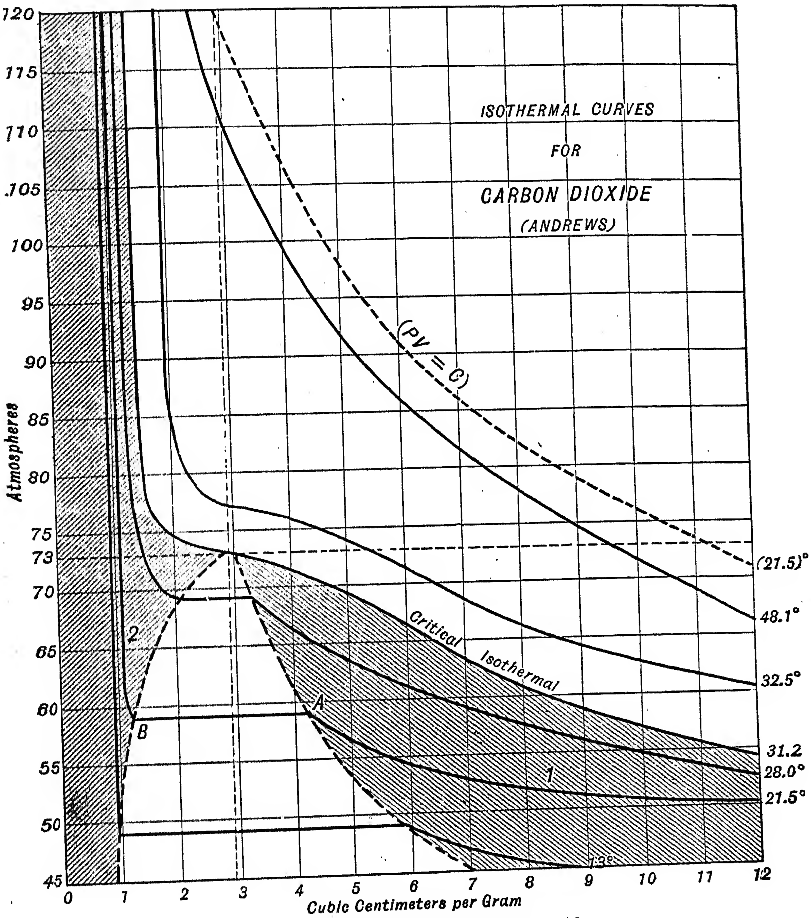


FIG. 201.—Isothermals of carbon dioxide.

liquid with a visible meniscus between becomes less as higher temperatures are chosen, until, when a temperature of 31.2° is reached, the horizontal part of the isothermal has disappeared, and no separation into liquid and vapor can be noticed during compression. The critical temperature is 31.2°C., and it may

be defined in another way as *that temperature above which it is impossible to liquefy a gas by pressure alone,* "to liquefy" meaning to cause a separation of the two states.

The line AB represents those conditions of pressure and volume in which liquid and vapor can exist in equilibrium with each other at the temperature of 21.5°, and a similar meaning attaches to the horizontal portions of the other isothermals. The dotted line through the ends of these horizontal lines surrounds an area representing all the physical conditions under which the liquid and its vapor can be in equilibrium with each other, and the highest point of this curve is the critical point, the coordinates of which are the *critical volume* and *critical pressure,* V_c and P_c, corresponding to the critical temperature t_c.

TABLE 15
CRITICAL DATA

Substance	Critical temperature °C.	Critical pressure (Atmos.)
Air	−140	37
Alcohol (ethyl)	243	63
Ammonia	132	111
Argon	−122	48
Carbon dioxide	31	73
Chlorine	144	76
Helium	−267.9	2.3
Hydrochloric acid	51.4	82
Hydrogen	−239.9	13
Nitrogen	−146	33
Oxygen	−118	50
Radium emanation	104.5	62.5
Water	374	218

That part of the dotted curve to the right of the critical point is called the "saturation curve." It evidently represents all possible conditions of saturated vapor, and, since the diagram is drawn for unit mass, the abscissæ of the points of this curve are the specific volumes of the saturated vapor at various temperatures. The left branch of the dotted curve is called the "liquid curve," and the abscissæ of this portion are the specific volumes of the liquid at the same temperatures. It is very clear, then, that the specific volumes of liquid and saturated vapor become equal at the critical point.

Above the critical temperature the distinction between liquid and vapor disappears, and the substance passes continuously and homogeneously from a rare, easily compressible condition, which we would call gaseous, to a dense, almost incompressible condition, which we would naturally call liquid. It is possible, by properly varying the pressure, volume and temperature, to pass from any condition 1 to any condition 2 without crossing the dotted curve, that is, without having the liquid distinct from the vapor at any time. This property is called the "continuity of state."

It is generally agreed to call a substance a vapor if its condition is represented by any point below the critical isothermal and to the right of the saturation curve, and a gas if it is represented by a point above the critical isothermal, though this distinction is not important. The properties represented by this set of isothermal curves for CO_2 are characteristic of all substances which have been studied.

319. Equations of State.—Many attempts have been made to derive equations for the isothermal curves of Fig. 201, corresponding to the equation $PV = RT$, which holds approximately for conditions far removed from the critical, but none have been entirely successful. One of the most satisfactory of such "equations of state" is that of van der Waals;

$$\left(P + \frac{a}{V^2}\right)(V - b) = RT$$

in which a, b, and R are constants for a given substance. This agrees fairly well with the results of experiment, though, instead of the straight portion AB of the isothermal (Fig. 201), the equation gives a curve which cuts the straight line in three points as shown in Fig. 202. The possibility of a continuous passage such as $DCBA$, below the critical temperature, from the vapor to the liquid condition, was suggested by James Thomson shortly after Andrew's work; but, except for portions of the curve from A toward B (undercooling a vapor free from nuclei) and from D toward C (superheating a liquid), it has not been realized experimentally and indeed seems quite unrealizable, since it would represent states in which an increase in volume would accompany an increase in pressure.

Corresponding States.—It was suggested by van der Waals that if the pressure, volume, and temperature were expressed in terms of the critical constants, P_c, V_c, t_c, for each substance, as units, instead of in atmospheres, cubic centimeters and degrees centigrade, for example, the "equation of state" would be the same for all substances, containing no constants peculiar to any one material. The states of all substances would *correspond* when they were represented by the same "reduced" values of P, V, and t. While this "theorem of corresponding states" is a necessary consequence of van der Waals' equation,

and may be safely and usefully applied between related substances, it is not true generally.

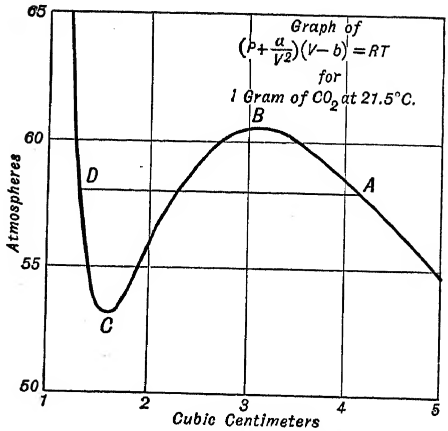


FIG. 202.—Graph of van der Waal's equation.

320. Thermodynamic Surface.—If the isothermals of Fig. 201 are placed in their proper position along the temperature axis, a smooth surface drawn through them forms the thermodynamic surface shown in Fig. 203, every

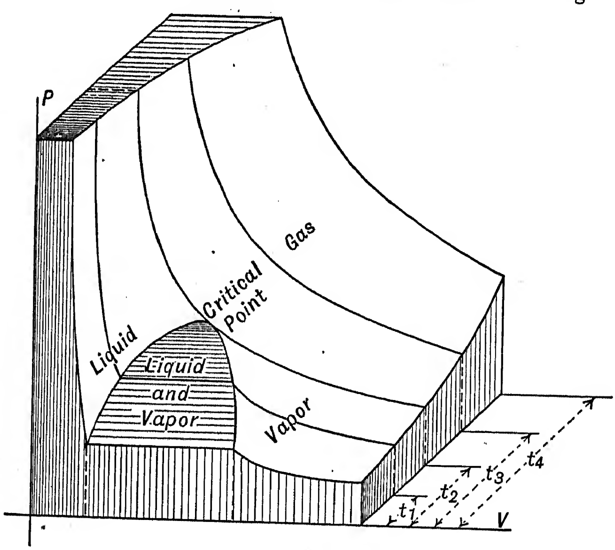


FIG. 203.—Thermodynamic surface, coördinates P, V, t.

point of which represents by its coördinates P, V, t, an equilibrium condition of 1 gram of a substance. The conditions under which the substance may exist

as liquid or gas or as a mixture of different states, are indicated on the diagram, and the triple point curve and the isothermals which we have already discussed are seen to be the projection on the Pt or PV plane of lines on this thermodynamic surface.

321. The Liquefaction of Gases.—By compression and cooling Faraday (beginning in 1823) liquefied carbon dioxide, sulphur dioxide, chlorine and several other gases not previously known in the liquid state. The temperatures he used were evidently below the critical temperatures as we now know them, but the problem was not thoroughly understood until the work of Andrews made it probable that extremely low temperatures as well as high pressures would be needed to liquefy oxygen, nitrogen, hydrogen and air, which, as late as 1877, were called *permanent* gases.

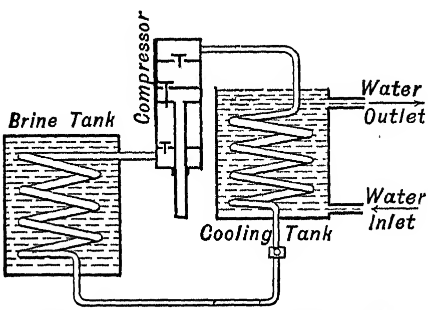


FIG. 204.—Ammonia refrigerating process.

The problem has been, then, one of devising methods of obtaining extremely low temperatures, and it has been so successfully solved that all known gases have now been liquefied. The following methods are used for obtaining low temperatures:

1. *Chemical method*, or the use of freezing mixtures, that is, mixtures of substances which in dissolving or combining absorb heat and thus lower the temperature. The lowest temperature reached by this method, $-82°$C., has been obtained by mixing solid CO_2 and liquid SO_2. This method is not used in recent liquefying processes.

2. *Evaporation Method.*—Fig. 204 illustrates this method as applied to the ammonia refrigeration process. A compressor exhausts ammonia gas from above the liquid, compresses it, forces it through tubes cooled by running water where the heat of vaporization and of compression is taken out and it again becomes liquid, and back through a reducing valve into the evaporating chamber.

The evaporating chamber is surrounded by the material to be cooled (circulating brine in the ordinary refrigerating system), from which the heat necessary to vaporize the ammonia is absorbed. The ammonia therefore absorbs heat in the evaporation chamber and loses heat in the cooling coils.

A series of such circulating systems, containing, for example, SO_2 in the first and CO_2 in the second, arranged so that the cooling of the CO_2 is done by the evaporating of the SO_2 is called the *cascade method*, by which Pictet in 1877 liquefied oxygen. The oxygen was compressed to several hundred atmospheres pressure in a tube

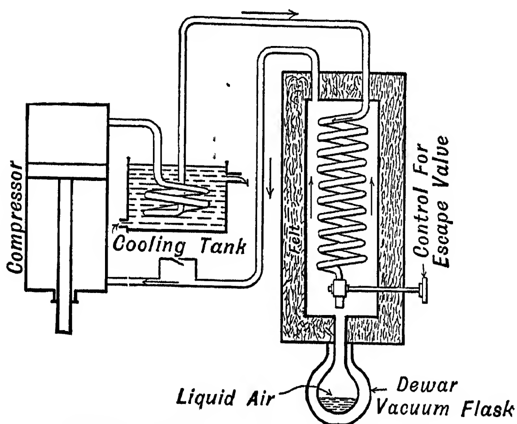


FIG. 205.—Linde's apparatus for liquefying air.

surrounded by the evaporating CO_2 and thus cooled to $-140°C$. Reference to Table 15 shows that at this temperature and pressure the oxygen would be liquid. Upon opening a cock the oxygen escaped in a white stream, indicating the presence of the liquid or solid By adding lower steps to the cascade it is possible to obtain very much lower temperatures, so that oxygen, nitrogen and hydrogen may be obtained liquid at atmospheric pressure.

3. *Regenerative Unbalanced-expansion Method.*—We have already considered (§296) the cooling experienced by all gases except hydrogen when they are forced through a small opening. This cooling is for air only about 0.25°C. per atmosphere decrease of pressure, but it increases with decreasing temperature. Hence if the cooled expanded gas is led back around the in-flowing compressed gas, as in Fig. 205, so as to cool it, the temperature at which

expansion takes place will be gradually lowered, until finally some of the gas will liquefy on expansion.

Apparatus for applying this principle was independently invented by Linde, Hampson and Tripler, and this method is now extensively used, commercially as well as scientifically, for liquefying oxygen, nitrogen and hydrogen. The expenditure in the work of compression of 10 h.p. for one hour will, by this process, produce from 2 to 4 liters of liquid air, the initial pressure being from 120 to 200 atmospheres. The heating which is observed with hydrogen at ordinary temperatures becomes zero at −80.5°C., as shown by Olzewski, and below that there is a cooling, so that, by initially cooling hydrogen below −80.5°C., it can be liquefied by the unbalanced expansion process, as was first done by Dewar in 1898.

TABLE 16

Boiling-point of Different Substances under Atmospheric Pressure, and Temperature Obtained by Boiling under Reduced Pressure

Substance	Boiling-point (atm. pressure)	Pressure (in mm. of mercury)	Boiling-point (reduced pressure)
Argon	−185.7	300	−194.2°C.
Carbon dioxide	− 78.2	2.5	−130
Helium	−268.9	10	−270
Hydrogen	−252	.....	−256
Neon	−245.9	2.4	−257.5
Nitrogen	−195.8	86	−210.6
Oxygen	−183	200	−194
Radium emanation	− 62	9	−127

4. *Regenerative Balanced-expansion Method.*—The first step in this method is the compression of the gas to about 40 atmospheres pressure, and the partial cooling of it by an "interchanger," analogous to the one used in the Linde process. The compressed and cooled gas is then admitted to a cylinder and allowed to expand against a piston, thus doing external and internal work, and being still further cooled to −160°C. or lower. The expanded gas is then led around the outside of a liquefying vessel containing air at 40 atmospheres pressure, and it cools sufficiently to liquefy some of the air. The expanded gas, after absorbing heat from the liquefying vessel, is led back through the interchanger to the compressor.

The advantage of this method over those of the Linde type lies in the greater amount of *external work* which the gas does, resulting in greater cooling. Moreover, being done against a piston, this work can be utilized. By combining three stages of expansion similar to the above, Claude has produced liquid air at the rate of 9 liters per 10 h.p. per hour.

In 1907 helium, the last gas to resist liquefaction, was liquefied by Kammerlingh Onnes by the unbalanced-expansion method, its boiling-point under one atmosphere pressure being −268.9°C. The lowest temperature attained by evaporating helium under reduced pressure is about −272°C. or 1° above absolute zero. These extreme temperatures are measured either by a gas thermometer containing helium at reduced pressure, or by a thermoelectric or resistance thermometer. The lowest temperatures which have been attained (about 0.1°K.) have been produced by magnetizing some of the paramagnetic rare earths at very low temperatures and allowing them to demagnetize and absorb energy. These low temperatures are measured by extrapolating curves connecting certain magnetic properties of the material with temperature.

CONVECTION OF HEAT

322. Convection, Conduction and Radiation.—Heat is transferred by three very different processes.

Convection is the transport of heat by moving matter, as, for example, by the hot air which can be felt rising from a hot stove.

Conduction is the flow of heat through and by means of matter unaccompanied by any motion of the matter, for example, the passage of heat along an iron bar one end of which is held in a fire.

Radiation is the passage of heat through space without the necessary presence of matter, for example, the passage of heat through the vacuum in the bulb of an incandescent lamp.

323. Convection.—Convection occurs in liquids and gases and is due to the change in density produced by rise in temperature. A volume of liquid or gas which varies in density in different parts is only in stable equilibrium when the densest portions are at the bottom and there is a regular decrease in density towards the top. Since (with the exception of water below 4°C., §278), liquids and gases expand on heating, thus diminishing in density, the heated portion will *rise,* and there will be an upward convection current of

hot substance and a downward convection current of cold substance to take its place. If heat is added at the *top* of an enclosed liquid or gas, there will be no convection, (except with water below 4°C.).

Common examples of *convection by liquids* are the distribution of heat through liquids heated from the bottom, as in the case of water in a tea kettle, and the distribution of heat through a house by the hot water system of heating, Fig. 206. The water is heated in A, rises through B, is cooled in the radiator and falls through C. On a large scale the Gulf Stream, Japan current, and other warm surface ocean currents which start near the equator are, in part at least, caused by convection, the return being a cold current flowing toward the equator along the ocean bed.

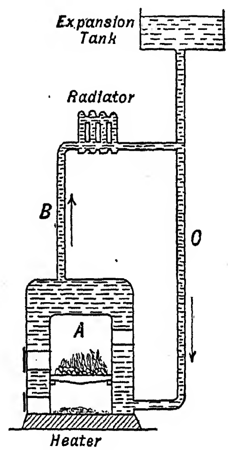


FIG. 206.—Transfer of heat by convection of hot water.

The hot-air furnace system of heating houses is based on *convection by gases*, the hot air rising from the furnace through the pipes and registers, and the supply of cold air coming usually from outside. The working of such a system can sometimes be improved by establishing a direct return from the coldest part of the house to the furnace, thus completing the indoor circulation.

"Natural ventilation" is also a convection process, an outlet being provided at the top of a room for the warm stale air, and an inlet at the bottom for the cold fresh air. The natural draft in chimneys has a similar cause; the higher the chimney the larger is the undisturbed column of warm air and hence the greater the draft. The mixing of currents of hot and cold air usually causes a flickering or "boiling" of objects seen through them, because light travels differently in hot and cold air. This effect may be seen by looking across a flat country in the hot sunshine, or over a hot pavement or stove.

The winds are largely convection effects, the simplest example being the "land" and "sea" breezes, which ordinarily blow from the sea to the land in the morning and from the land to the sea at night. These are the return currents which replace warm air which rises from the quickly heated land in the morning and from the warmer, more slowly cooling sea at night.

CONDUCTION OF HEAT

324. Characteristics of Heat Conduction.—As we have said, conduction is the flow or passage of heat energy through and by means of matter unaccompanied by any obvious motion of matter, as, for example, the passage of heat through the bottom of a kettle to the water inside.

The direction in which heat will flow between two points, whether from A to B or B to A, is found to depend on the relative temperatures of A and B, heat always flowing from the point at higher temperature to the one at lower temperature. The greater the difference of temperature between two points, other conditions being the same, the more heat will flow per second, for which reason a kettle boils more quickly over a hot fire than over a low one. But, with a given temperature difference between the fire and the water, boiling will take place more quickly with a thin-bottomed kettle than with a thick. The temperature difference between two points, A and B, divided by the distance, l, between them, or $\frac{t_A - t_B}{l}$, which is the average fall in temperature per centimeter between A and B, is called the *temperature gradient.* The above statements of the dependence of conduction on temperature difference and distance may be combined by saying that **the amount of heat conducted per second between two points is directly proportional to the temperature gradient.** To pursue that same example, common experience dictates that the kettle should have a broad bottom in contact with the stove. This is an illustration of the fact that, other conditions being the same, the amount of heat conducted per second is *directly proportional to the area* through which it can flow.

Finally, the rate of flow of heat, other conditions being the same, depends greatly upon the material through which it must flow, substances being roughly divisible into "good conductors," which permit under given conditions a large flow of heat, and which in general are metallic, and "poor conductors," which permit a small flow of heat and are in general non-metallic, such as wood, glass, asbestos, leather, linen. Examples of this difference are very common. A glass of hot water may be handled, while a metal cup containing the same water will be too hot to touch. Handles of heating vessels are made of wood, or, if of

metal, are covered with string or cloth, so that they may be touched. Given two bodies, one metal and one wood, at the same temperature, below that of the hand, the metal one feels much cooler because the heat it takes from the hand quickly spreads through the mass, while, with the poorly conducting wood, the heat remains near the surface of contact, which quickly rises in temperature. Thus the wood feels warmer because (after the first instant) it *is* warmer, where it is touched. The rate of heat flow through a body will depend not only on the substance composing it but on its condition of subdivision and density. Thus saw-dust conducts less readily than wood, and the small conduction through cork is partly due to the reduction of effective cross-section by air holes. Also substances when moist conduct better than when dry, because water, which fills the pores, is a better conductor than air.

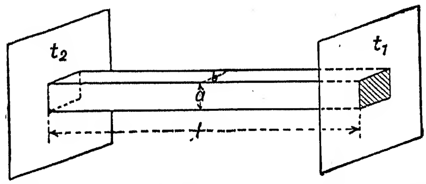


Fig. 207.—Illustrating simple case of conduction of heat.

The characteristic of bodies which determines the rate of flow of heat through them is called their *thermal conductivity,* and the numerical measure of this characteristic, is called the *coefficient of thermal conductivity,* which will be defined in the next paragraph.

325. The Coefficient of Thermal Conductivity.—In order to group together the previous statements and obtain an exact definition of the conductivity coefficient, consider the passage of heat from the region (2), Fig. 207, at the uniform constant temperature t_2, to (1) at the temperature t_1, by conduction along the rectangular bar of cross-section A $(= ab)$ and length l, no heat being allowed to escape from the sides of the bar. If H is the heat which passes in a time T, then from the statements above, it follows that, for a given substance,

$$H \propto \frac{T(t_2 - t_1)A}{l} \tag{1}$$

and, by introducing a proportionality factor K, properly chosen for each substance, this may be written as an equation,

$$H = \frac{KT(t_2 - t_1)A}{l}. \tag{2}$$

K is the coefficient of thermal conductivity of the material. If, as a special case, we take $A = 1$, $T = 1$, $t_2 - t_1 = 1°$, $l = 1$, then $K = H$, or K for a given substance is **the heat flow per unit time per unit area with unit uniform temperature-gradient.** In the c.g.s. system, the units would of course be the centimeter, second, centigrade-degree and calorie. This statement defines K at a definite mean temperature, $(t_1 + \frac{1}{2})°$, but, since K varies with the temperature, equation (2) will not be true for any great difference of temperature, $t_2 - t_1$, unless K stands for the *mean value* between these limits, and the temperature gradient is *uniform.*

Equation (2) is the basis of most methods for measuring K, but while they are very simple in principle they are very difficult to carry out accurately.

If a uniform bar of the substance has its end faces maintained at fixed temperatures, for instance $t_2 = 100°$C., and $t_1 = 0°$C., by contact with steam and melting ice respectively, its sides being protected from loss of heat, and if we find, by measuring the amount of ice melted, the heat which flows through the bar in a time T, and also the length and cross-section of the bar, K may be computed at once. This will, of course, be the mean value of K between 0° and 100°C. Other calorimetric methods are usually employed for measuring the amount of heat flowing in a given time, and with poor conductors a slab or plate of the substance rather than a rod must be used to give a measurable rate of flow.

The average temperature gradient in the earth's crust is about 1°C. rise per 100 ft. of descent below the surface. This means a continual loss of heat from the interior, small in amount, however, on account of the low conductivity of rocks and soil. For the same reason (low conductivity) the daily variations of surface temperature penetrate only about 2 ft., the annual about 30 ft. (§239). The freezing of a plane surface of thin ice, such as forms on the surface of a lake, is an interesting example of the influence of conductivity, the thickness increasing as the *square root* of the time.

326. Conduction in Liquids and Gases.—In order to measure conduction in liquids, as distinct from convection, heat must, of course, be added at the *top* (except with water between 0° and 4°C.), since, if heated from below, the warm expanded liquid would rise

while, if heated from above, it will remain in place. Since many liquids are transparent to radiation (§330) there is also danger that we may confuse radiation across the liquid with conduction through it. The conductivity of liquids is, in general, about that of solids of low conductivity, except in the cases of mercury, which is metallic and a good conductor, and water and some aqueous salt solutions, which are intermediate between the metallic and the non-metallic solid conductors.

The masking of conduction by convection and radiation is even more likely to occur in gases, because of their greater mobility, greater transparency, and lower *real* conductivity. The conductivity of hydrogen and helium is much greater than that of other gases. This follows naturally from the kinetic theory, since they have the smallest molecular weights, therefore the highest molecular velocities at a given temperature, and therefore hand on kinetic energy from molecule to molecule with the greatest rapidity. The conductivity of gases is, through wide ranges, independent of the pressure, as theory also indicates should be the case; it decreases, however for very low pressures. This means that the hollow walls of a Dewar flask or thermos bottle must be *highly* evacuated to give good insulation. On account of their extremely low conductivity, air layers enclosed in or between solids, such as air spaces in house walls and in the walls of refrigerators, in pores in cloth and in fur or feathers, are chiefly responsible for the low conductivity found in these cases.

327. Conductivity of Alloys and Crystals.—The conductivity of copper is increased by compression, that of steel diminished by hardening. The conductivity of alloys is not, in general, simply proportional to the relative amounts of the pure metals forming the alloy, but may have decided minimum values in case compounds are formed. In non-isotropic solids, such as wood and crystals, the conductivity depends upon the direction of flow, being in the case of wood two or three times as great along the fiber as at right angles to it. In crystals the axes of symmetry for heat conduction coincide with the crystalline axes, and the conductivity is different in different directions, as may be very prettily shown by means of a thin plate of crystal coated with wax on one side, and having a wire passing normally through the center. If the wire is carefully heated the wax will gradually melt and the limit of melting will be, in general, an ellipse and not a circle, as it would be with an isotropic plate. The marked decrease in effective conductivity resulting from breaking up a solid has been referred to, but this, as well as the effect of compression on a substance like felt or cotton, is not a change in the property of the substance itself, but merely a change in the amount of poorly conducting material (air) mixed with it.

328. Other Problems of Heat Conduction.—The mathematical theory of heat conduction dates back over a century to Fourier. The case discussed in §325 of the steady flow of heat between two surfaces at constant temperature is a simple and common one, but there are many others. Some of them have to do with temperatures in the earth such as occur in the penetration of "cold waves" and "heat waves." These diminish rapidly in amplitude as they travel down (Fig. 139). Water pipes at a depth of 6 feet might encounter a maximum temperature variation of 18°; *i.e.*, 9° above or below the mean. If the mean temperature was 9°C. or above there would be no danger of freezing. Latent heat considerations in connection with the freezing of the soil would give additional protection.

TABLE 17

THERMAL CONDUCTIVITIES

Substance	Conductivity	Substance	Conductivity
Aluminum	0.485	Lead	0.083
Brass	0.260	Nickel	0.142
Air	0.00005	Oak	0.0006
Concrete	0.0022	Platinum	0.166
Copper	0.927	Porcelain (Berlin)	0.0025
Cork	0.00013	Quartz ‖-axis	0.030
Cotton wool	0.0001	Quartz, ⊥-axis	0.016
Earth's crust	0.004	Sawdust	0.00012
Flannel	0.00023	Silk	0.0002
Glass	0.0024	Silver	1.002
Gold	0.700	Tin	0.155
Ice	0.005	Water	0.0014
Iron	0.144	Zinc	0.265

The existence of an average temperature gradient for the earth's crust gives rise to a continual flow of heat from the hot interior to the cool surface, but the total amount in a year would only be sufficient to melt a layer of ice a little over half a centimeter thick. Calculations based on the time it would take to establish this gradient, starting with an earth at an assumed initial high temperature, form one method of estimating the age of the earth, that is, the time since the present crust was formed on the planet. Such calculations give a minimum value of the order of 100 million years, which is much smaller than the age now accepted. This is 2 or 3 billion years, and has been computed from certain radioactive disintegration phenomena found in very old rocks (§550).

If the boundary surface of a large cold solid is heated to and kept at a higher temperature, heat will flow in, and the time required for points inside to reach a given temperature will be approximately proportional to the square of the distance from the surface. This has a bearing on the design of fireproof safes and fireproof walls, for doubling the thickness will give about four times the protection as regards the delay in penetration of disastrously high temperatures. As already remarked (§325), somewhat the same sort of law holds in the freez-

ing of a plane surface of ice on a lake. Here the time required for the freezing of a given thickness increases as the square of the thickness.

329. The Nature of Conduction.—Since we have agreed that heat energy is in large part kinetic energy of motion of molecules, atoms, and electrons, it is natural to think of this motion (heat) as spreading through a substance by collision of these particles with each other. Heat added to one side of a body will increase the average energy of motion of the molecules on that side, and will be gradually handed on by impact to the slower moving ones and so will spread through the mass, much as a disturbance originating at one point would spread through a closely packed crowd by repeated pushing and jostling.

While this has been the common idea of the nature of the process, J. J. Thomson and others have attempted to account for the matter in an entirely different way, namely, by the *convection* of "free" electrons as defined in §367. According to this hypothesis the addition of heat to one part of a substance increases the kinetic energy of the free electrons in this region, and there results, not only the transfer of energy by impact, but the actual diffusion of fast moving electrons from the hot to the cold region. The motion of these electrons, according to this hypothesis, also constitutes an electric current, so that this explanation of heat conductivity would very easily account for the remarkable observed fact that those substances which conduct heat readily, such as metals, also conduct electricity readily, the electrical and thermal conductivities being in a fairly constant ratio for most metals at ordinary temperatures. This "free electron" picture of the process of conduction has so far proved to be too simple to account for all the observed facts, especially the very complicated changes in electrical conductivity produced by high pressures and extreme low temperatures (super-conductivity). Another point of view, more hopeful but more complicated, connects thermal conduction with the elastic properties of solids, but the details cannot be considered here.

RADIATION

330. Radiant Energy.—Radiation is the process by which energy is transmitted through space without the necessary presence of matter. While being transmitted in this way energy is called *radiant* energy and is not heat, since the latter is energy in a particular relation to matter. That energy may pass through matter and still not be heat may be shown by allowing the sun's rays to pass through glass and fall upon a blackened thermometer, which may be very decidedly heated, though the glass remains cool. Radiation differs most strikingly from convection and conduction in *speed*. Time a convection current (by means of smoke or dust) and the velocity will usually not be many feet per second; thrust one end of a silver rod into hot water and it will be several seconds before a noticeable effect can be felt a few centimeters above the surface; but an opaque screen for cutting

off radiation produces a practically instantaneous effect even at a great distance.

The early idea regarding the nature of radiation was that it was a streaming of fine particles or "corpuscles"—that is, a convection. It is now usually considered to be a *wave disturbance*, such as has been discussed in Wave Motion, analogous to the waves which travel over a water surface. The "disturbance" of which the water waves consist is an up-and-down motion of the water particles, and this disturbance travels *forward* while the water moves up and down. The "disturbance" in a radiation wave is a transverse electric and magnetic force, which changes in direction and amount as a wave passes a point (just as the motion of the water particles changes from up to down), and which would move a compass needle if we could make one small enough for it to act on. The characteristics of a radiation wave are: the *period;* the *wave length;* and the magnitude of the electric force which a wave produces as it passes, or the *amplitude* (corresponding to the height of a "crest" in a water wave) which determines how strong the wave is, how much energy it represents, and is quite independent of the wave length. A strong wave may be long or short, a long wave weak or strong.

Radiation travels in free space with the same speed as light (§617), and like light it can be reflected by mirrors and refracted by lenses and prisms. These and other facts prove conclusively that radiation waves are of exactly the same nature as light waves, in fact that *light* consists simply of those radiation waves whose lengths lie between 0.0004 and 0.00076 millimeter and which affect the eye. These waves lie near one end of the entire known range of radiation wave-lengths, which is from about 1×10^{-8} mm. to nearly 1 mm. and is called the *radiation spectrum.* Of this the *visible spectrum* evidently forms but a very small part. That part of the spectrum called the "infra-red," of longer wave-length than the visible, contains usually the waves of greater energy, the most important for the radiation of heat, and these waves were formerly called "radiant heat." Radiation waves can travel through *free space,* their transmission being one of the fundamental properties of space as we know it.

The wave idea of radiation appears to be inconsistent with certain observations (Compton effect) with X-rays, which, according to the wave theory, would be waves of very short length, of the order

of 1 to 10 × 10^{-8} mm. To summarize briefly, one may say that radiation behaves in some respects (where it falls on matter) as if it were constituted of lumps or "quanta" of energy, while in other respects it seems to consist of waves. The present point of view is therefore a combination of the wave theory with a sort of corpuscular theory.

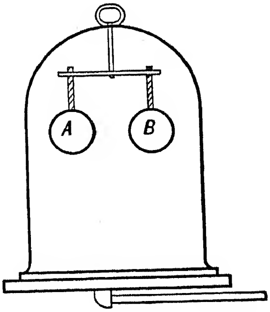


FIG. 208.—Radiating bodies in a vacuum.

331. Law of Exchanges.—If two bodies at different temperatures, for example the copper balls A and B, Fig. 208, are put in a vacuum and not in contact, equality of temperature will be established by radiation, the hotter body A on the whole radiating heat to the colder B. If, without changing the temperature or condition of B, A is cooled till it is the colder of the two, the net exchange of heat by radiation will now be from B to A. Since we have not altered B in any way, we conclude that B was radiating to A in the first case also, but that A was radiating more to B. Radiation is then always a reciprocal process, or one of exchange. This is the *Prevost law of exchanges*, according to which radiation equilibrium is the result of equal streams of radiant energy in opposite directions, and does not indicate the cessation of radiation.

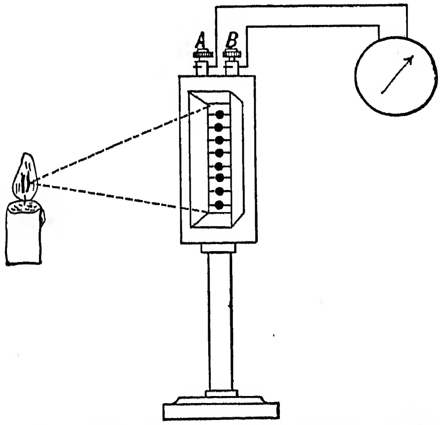
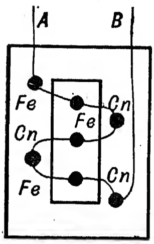


FIG. 209.—Thermopile for measuring radiant energy.

332. Measurement of Radiation.—In order to measure radiation it is converted into heat by absorption in matter, and the heat

is then measured by the temperature change which it produces. To make this process sufficiently sensitive, recourse is had to the thermo-electric or resistance methods for measuring temperature described in §§268, 269.

The *thermopile*, Figs. 209, 210, consists of one or more "junctions" of different metals, iron and constantan, or better two alloys of bismuth-antimony and antimony-cadmium, arranged as shown in Fig. 210, so that one set of similar junctions can be exposed to radiation while the other set is protected. To increase the amount of radiant energy intercepted, the exposed junctions should be covered with thin blackened silver or copper disks, and similar disks should be put on the other junctions. If the final elements are connected by wires to a very delicate galvanometer, very slight changes in temperature of one set of junctions, of the order of $\frac{1}{1,000,000}$ °C. or less, will produce a readable deflection, and will correspond to a very weak stream of radiant energy falling on the exposed junctions, such as, for example, the radiation from a single candle at a distance of 50 meters. To be quick-acting and sensitive the mass of the junctions should be small. One of the most sensitive types of thermopile consists of one or more junctions of fine bismuth and tellurium wires with thin blackened gold disks to absorb the radiation. It is used in a high vacuum, which increases the sensitiveness many fold, because loss of heat from the junctions is thereby diminished, and they are accordingly warmed more by the radiation.



Detail of Junctions

FIG. 210.—Detail showing arrangement of the exposed (inner) and protected (outer) junctions of a thermopile.

The bolometer, Fig. 211, is also a sensitive instrument. It consists essentially of two similar strips of very thin (0.001 mm.) blackened platinum mounted side by side, having exactly the same resistance, and arranged in a Wheatstone bridge (§453), so that any unequal changes in resistance of the strips can be very sensitively measured. If one strip is exposed to radiation, its temperature and hence its resistance will change.

Another very sensitive instrument which may be used either to detect or to measure radiation is the *Crookes' radiometer*, the essential feature of which is a very light vertical shaft carrying several flat vertical vanes, like paddles, of aluminum or glass mounted so

that their planes intersect in the shaft. One set of corresponding faces of the vanes are blackened, the other set is polished, and the whole is very delicately pivoted or suspended inside a vessel of clear glass in which the air pressure has been reduced to about $\frac{1}{100}$ of a mm. of mercury. If the vessel is placed in sunlight or other sufficiently intense beam of radiation the shaft rotates in such a sense that the exposed black surfaces move *away from* the incident beam. While the action is a complicated one, the primary cause is the greater absorption of radiation by the blackened surfaces. Being warmer, the molecules of the black surfaces are in more violent motion than those of the bright surface, and are thus able to "kick

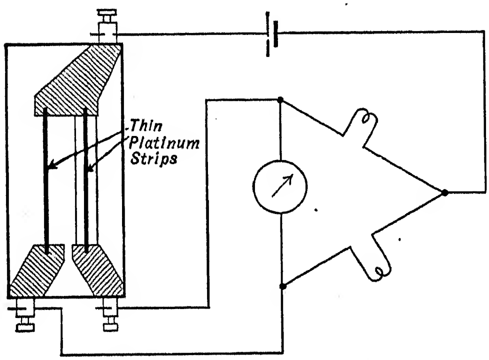


FIG. 211.—Bolometer for measuring radiant energy, and Wheatstone's bridge for measuring change of its resistance.

off" the impinging air molecules more violently. This instrument has been refined and improved for purposes of accurate measurement by Nichols and Pringsheim.

333. Emission, Absorption and Reflection.—*Emission* is the starting of radiation waves. The *conversion of the energy of a wave into heat by passage through matter is called absorption.* A substance is opaque to radiation when it will not allow the radiation to pass through, as, for example, wood and metals are opaque to light. Absorption by a very thin surface layer of a strongly absorbing substance is called *surface absorption.* Such a surface, if polished, will also, in general, reflect very well.

The *absorbing power* of a surface is the ratio of the radiant energy absorbed by the surface to the amount incident upon it. The absorbing power of a given surface is, in general, different for different wave-lengths of radiation. Let A_λ be the absorbing

power for the wave-length λ and A the total absorbing power for all wave-lengths. The values of A and A_λ are practically independent of temperature. The *emissivity* of a surface is the total radiant energy, in ergs, which the surface sends out per square centimeter per second, this radiation being caused by the heat of the surface. The hotter a body the more it radiates, that is, emissivity increases with temperature. We shall denote the *total emissivity* for all wave-lengths by E, and the emissivity for the wave-length λ, or *partial emissivity*, by E_λ.

The *reflecting power* of a surface is the ratio of the radiant energy reflected from the surface to the energy incident upon it. The reflecting power of a surface is different for different substances and for different wave-lengths of radiation. Thus a polished silver surface will reflect about 82 per cent. of all blue light falling upon it, about 92 per cent. of incident yellow light, and about 98 per cent. of energy in the form of long infra-red waves, while a polished iron surface reflects about 57 per cent. of yellow light and from 78 to 97 per cent. of the energy of infra-red waves.

That there is a close connection between the absorbing power and emissivity of a surface can be shown, for example, by heating a bit of white china with blue markings, which look *dark* against the light china at ordinary temperatures because they *absorb* more light, but look bright against the china at high temperatures, showing that they *emit* more light. Similarly black ink marks on platinum look bright when heated. *In general good absorbers are good radiators.* That this must be so follows from a consideration of a body B suspended inside an exhausted opaque vessel C. Experience shows that B and C will come to the same temperature by interchange of radiation, and when equilibrium is reached B must absorb per second as much as it radiates. Hence if B is a good absorber it must be a good radiator, and vice versa.

In the case of opaque surfaces, all incident radiation which is not reflected will be absorbed. Hence if I is the incident radiation, we must have $I = AI + RI$, or $A + R = 1$; similarly $A_\lambda + R_\lambda = 1$. It has been found for a number of metals that at high temperatures the reflecting power varies, and in a different way for different wave-lengths.

334. Kirchhoff's Law.—The exact relation between absorbing power and emissivity, deduced theoretically by Kirchhoff and called after him *Kirchhoff's*

Law, is that *the ratio of the emissivity to the absorbing power is the same for all surfaces at any one temperature,* or

$$\frac{E}{A} = \mathrm{E}$$

and similarly, as regards any particular wave-length

$$\frac{E_\lambda}{A_\lambda} = \mathrm{E}_\lambda$$

where E and E_λ are constants independent of the substances.

335. A Perfect Absorber and Perfect Radiator.—If we could have a surface which *absorbed all the radiation falling upon it,* called a *perfect absorber* or *black body,* then for this surface

$$A = A_\lambda = 1, \text{ and consequently } E = \mathrm{E} \text{ and } E_\lambda = \mathrm{E}_\lambda$$

In other words, the constants E and E_λ are the total and partial emissivities of a black body. Since A and A_λ can never be greater than 1, it follows that a *black body* has the *greatest possible* total and partial *emissivity,* at any temperature, and it is, therefore, also called a *perfect radiator.* A hollow opaque body having a *small* opening in the walls is a very close practical approximation to a black body, because radiation entering through the opening is partially reflected and re-reflected inside and thus eventually almost all absorbed. Also a sharp conical hollow or wedge-shaped cleft with straight opaque polished sides, no matter of what they are made, absorbs all radiation entering it. Conversely, if the walls of the enclosure or cone or cleft are uniformly heated, the radiation which leaves the opening will be that of a *perfect radiator* at the temperature of the walls, since, as we concluded in §334, it must be independent of the nature of the enclosure. These are all practicable ways of realizing a *perfect absorber* and *perfect radiator.*

336. Total Radiation and Temperature.—The radiation of all bodies increases with the temperature, but the laws governing this increase are not as yet known except for a perfect radiator, for which Boltzmann deduced in 1883 the law previously suggested by Stefan, that

$$\mathrm{E} = sT^4,$$

T being the absolute temperature of the surface and s a constant which later work has shown to be approximately 5.7×10^{-5} ergs per square centimeter per second. According to this law the radiation from one square centimeter of black body surface at 400° absolute (127°C.) would heat one gram of water 1.5°C. per minute. If one black body surface at temperature T is radiating to another surrounding it at temperature T_1, then the net or differential radiating power will be, from the law of exchanges,

$$E = s(T^4 - T_1^4)$$

While this law can be deduced only for a black body, it is found to hold approximately for other surfaces. For the ordinary metals, however, the exponent of T is not 4 but varies from 4.5 to 5.5. Rewriting the equation in the form

$$E = s(T - T_1)(T^3 + T^2T_1 + TT_1^2 + T_1^3),$$

it is evident that, if T_1 does not differ much from T, we have approximately

$$E = 4sT^3(T - T_1) = K(T - T_1) \text{ (for } T_1 \text{ constant)}$$

A similar relation, known as *Newton's law of cooling*, is found to hold for the loss of heat by *combined radiation* and *convection*, and was enunciated by Newton as follows:

The heat lost by radiation and convection by one body to another surrounding it is proportional to the temperature difference between the two. This is a convenient relation to use and is quite accurate for small temperature differences.

337. Distribution of Energy in the Spectrum.—As the temperature of any radiating surface is raised, the energy emitted in every wave-length increases also, but not in equal proportion. It is a matter of common experience that the light emitted from a hot radiating surface changes in color as the temperature of the surface is raised, changing from red to yellowish, then to white and finally having a blue-white color at extremely high temperatures.

If, for a given surface, we plot the values of E_λ as ordinates, and the corresponding values of λ as abscissæ, for any one temperature, we obtain what is called the *energy curve* for this temperature. For example, the energy curves of Fig. 212 show the distribution of energy in the spectrum of a perfect radiator at several temperatures. Such curves have a general similarity for all surfaces, the emission being weak for short wave-lengths, rising to a maximum, and diminishing again for long wave-lengths. As the temperature of the radiating surface is increased, all the ordinates of the curve increase, and the maximum shifts toward the short wave-lengths. This shift of the energy curve, resulting in an increasing proportion of blue in the emitted light, accounts for the change in color of an incandescent body, which was just referred to. For a perfect radiator at 100°C. the maximum of

the energy curve lies at a wave-length of about 0.008 mm., while for carbon at the temperature of the arc it has shifted to the edge of the visible spectrum, and for the sun it is in the yellow.

The emission, E_λ for any wave-length for a black body is given with very great accuracy by the expression

$$\log E_\lambda = K_1 + K_2/T$$

where K_1 and K_2 are constants, different for different wave-lengths,

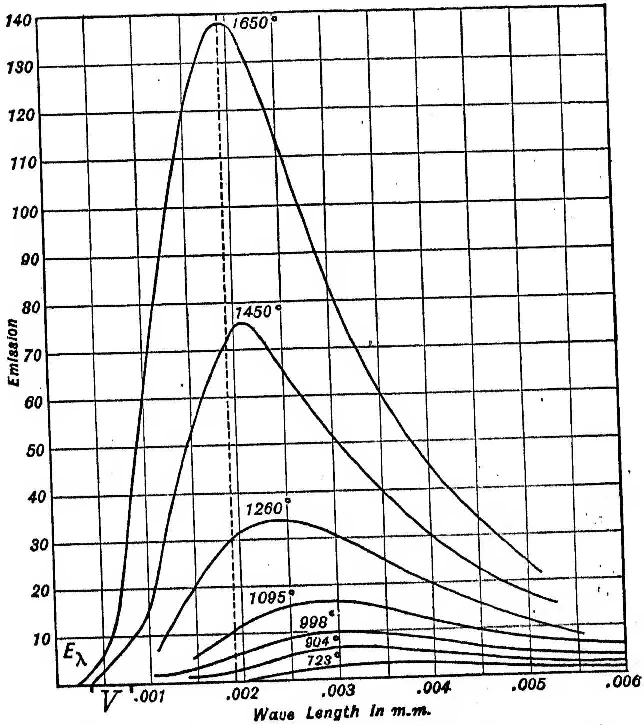


FIG. 212.—Curves showing distribution of energy in the spectrum of a perfect radiator at various temperatures. The extent of the visible spectrum is denoted by V. See also Fig. 601.

and T is the absolute temperature. For some other radiating surfaces E_λ has been found to follow quite closely the same law, though with different constants.

338. Radiation Pyrometry.—By this is meant the measurement of high temperatures by observing the variation with temperature of either total emissivity or partial emissivity. In the former case radiation of all wave-lengths from the surface whose temperature is to be measured is allowed to fall

on a thermopile of some sort, and the resulting deflection of a voltmeter or galvanometer is noted. By observations on a surface at known temperatures the instrument can be calibrated so as to indicate temperatures directly. Instruments of this sort are the Féry or Thwing total radiation pyrometers. If the instrument is *calibrated* by using a perfect radiator, and *used* on another surface, it will indicate, not the true temperature of this surface, but the temperature of a perfect radiator, which would radiate with the same total intensity as this surface. This is called the *total brightness temperature* of the surface, and will usually be lower (it cannot be higher) than the true temperature.

Optical pyrometry makes use of the partial emissivity. The method consists in comparing the radiation of a given wave-length (usually red) from the surface whose temperature is desired and from a comparison source, usually a small incandescent lamp. In using the instrument the electric current is measured which is required to heat the comparison lamp so that it disappears when viewed against the hot body. The instrument is calibrated by observations on a black body at known temperatures, and the radiation laws given by the equation in §337 may be used to extend the scale beyond the region of

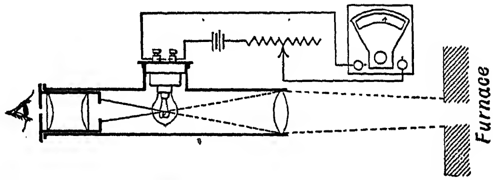


FIG. 213.—Optical pyrometer for high temperature measurements. Ammeter for measuring current in comparison lamp.

possible comparison. In this way measurements have been made up to 3600°C. If used on a surface other than a perfect radiator, it will give a "brightness temperature," less than the true temperature. Examples of this type of instrument are the Holborn, Morse, and Wanner optical pyrometers, the first named being shown in Fig. 213. Radiation pyrometry at present is the only satisfactory method available above about 1750°C.

THE CONSERVATION OF ENERGY

339. The Transformation of Mechanical Energy into Heat.—Since heat is energy and can be produced by the transformation of mechanical energy, it is of great importance to determine just how much mechanical energy is equal to a unit quantity of heat. In the c.g.s. system, **the mechanical equivalent of heat is the number of ergs equivalent to one calorie.** The symbol for it is J.

The first careful determination of this important quantity was made by Joule in 1843, before the caloric theory of heat was finally overthrown. Rowland in 1878 carried out one of the most reliable determinations of J that have been made, the method

being an improvement of one used by Joule many years before. A calorimeter (Fig. 214) contains water and one fixed and one movable set of paddles, the latter being driven by a shaft through the bottom of the calorimeter, and the former so arranged that the water can not rotate as a mass, but will be violently churned. The paddles are driven at a steady speed by a steam engine, and the calorimeter is prevented from rotating by the couple applied through two cords which pass tangentially from a carefully turned rim of radius R, and, after passing over frictionless pulleys, carry two weights of M grams each. The resisting couple experienced by the paddles in their motion through the water must be equal in

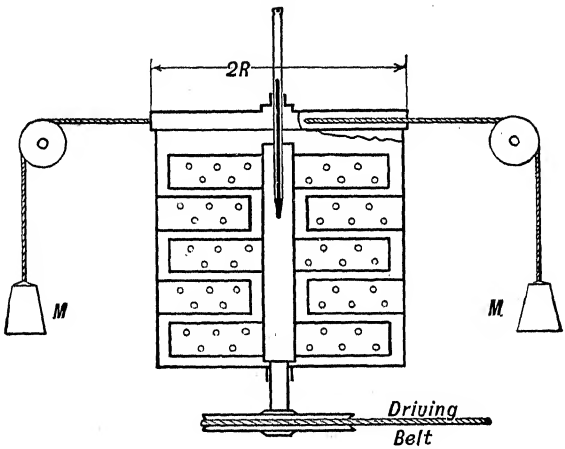


FIG. 214.—Apparatus for measuring the mechanical equivalent of heat.

magnitude to the couple which the water exerts tending to rotate the calorimeter. The resisting couple increases as the speed of rotation is increased. If for a given speed the weights M are so adjusted that the calorimeter does not rotate, and if we let L represent the moment of the resisting couple, then

$$L = 2RMg.$$

From this the work done by the paddles can be calculated.

$$\begin{aligned} \text{Work in one revolution} &= 2\pi L = 4\pi RMg, \\ \text{Work in } N \text{ revolutions} &= 4\pi NRMg \end{aligned}$$

If, after the N revolutions, the temperature of the calorimeter has risen $(t_2 - t_1)$ degrees and if m and m' are the mass of water contained

in the calorimeter, and its water equivalent, respectively, then $(m + m')(t_2 - t_1) =$ the heat added to and retained by the water. To this must be added the heat lost by convection and radiation, which we shall denote by H. There is no correction to be applied to the expression for the work, for only the work done against the frictional forces in the calorimeter is measured. Hence $4\pi NRMg$ ergs are equivalent to $(m + m')(t_2 - t_1) + H$ calories

or $$J = \frac{4\pi NRMg}{(m + m')(t_2 - t_1) + H} \text{ ergs per calorie}$$

Allowance must also be made for variation in the specific heat of water.

Other methods have been used in which the friction occurred between metal surfaces, the heat being absorbed from them by the water of the calorimeter or by a stream of water flowing past them (see §289). The latter arrangement is identical with one of the standard forms of *absorption dynamometers* used in engineering practice for obtaining the power developed by an engine or motor. Since the mechanical value of electrical work can be very accurately determined (§423), it is possible to determine J indirectly by converting electrical energy into heat, as has been done by Callendar and others. This method is simpler than the direct one. The best value is

$$J = 4.185 \times 10^7 \frac{\text{ergs}}{\text{cal}_{15}}$$

which is practically Rowland's value and is probably correct to $\frac{1}{30}$ per cent. The numerical value of J depends, of course, on the units, the following being also used

$$J = 427 \text{ kilogram-meters per large calorie.}$$
$$J = 778 \text{ foot-pounds per B.T.U.}$$

340. The Law of Conservation of Energy.—We have already become familiar in Mechanics with the transformation of kinetic energy ($\frac{1}{2}mv^2$) into potential energy when work is done against mechanical forces, and we have given reasons for believing that heat is a special form of kinetic and potential energy. Later we shall have to deal with electric and magnetic forces and work done against them, giving us the idea of electric and magnetic kinetic and potential energy, while chemistry deals with chemical

potential energy, though this is now believed to be electrical in nature.

After the growth of the idea that heat is energy, and Joule's early (1843) determination of J, Helmholtz, in 1847, formulated the idea that not only heat and mechanical energy, but all forms of energy are equivalent, and that a given amount of one form cannot be made to disappear without an equal amount appearing in some of the other forms. For example, when the potential energy of a wound clock spring disappears, heat, caused by work done against frictional forces, appears in the clock, while energy of sound waves and kinetic energy of motion of parts of the clock are also produced. Again, the heat energy of steam may be transformed into mechanical energy by a steam engine and given to a dynamo, which does work against electric and magnetic forces, producing some heat but largely electric potential energy, which in turn is changed, by the flow of an electric current, partly into heat in the wire, but largely into mechanical work by a motor, or into light and heat by an electric lamp. This idea of equivalence may be expressed in many ways, such as,

Energy is indestructible.

The total amount of energy in the universe is constant.

The energy required to change a system of bodies from one state to another state is independent of the particular intermediate states through which it passes.

These are all statements of the *Law of Conservation of Energy*, of which the last is perhaps the best, because we cannot deal with the universe, nor can we measure the *total amount* of energy present in any body. The fundamental idea is that all processes, such as the change of the energy of steam into mechanical energy and light above mentioned, consist in drawing a stream of energy from some source and then dividing and diverting that stream into various channels such as heat, mechanical work, light, etc. Common experience shows us that it is always very easy to convert any other form of energy into heat. Whenever a bell is rung by a battery, or a pump operated by a wind mill, *some* of the energy of the battery or the wind is changed into heat.

Like all the greatest fundamental physical laws, the law of conservation of energy is not capable of direct proof, but is an *assumption consistent with all known facts*, and is to be accepted until some phenomena are discovered with which it is inconsistent.

It is of the widest possible application and is the chief basis of all physical, astronomical and chemical reasoning, as well as of engineering practice. It leads us to exclude at once all "perpetual motion" devices, which purport to obtain mechanical work from nothing.

THERMODYNAMICS

341. First Law of Thermodynamics.—Thermodynamics is the analysis and discussion of the problems of converting heat into other forms of energy, and other forms of energy into heat, and consists in the deduction of consequences from two very general principles, the first one being the law of *conservation of energy* (§340).

Considering a body or a system or group of bodies as distinct from its surroundings, we have already (§262) defined the term "internal energy," for which we shall use the symbol U, as the entire energy which the system contains. As was pointed out earlier, we have no knowledge of the value of U in any case, but we can study the *changes* in U. If the reactions between the system and outside bodies are such as to permit the passage of heat to or from the system, and the doing of work on or by the system, then it follows from the law of conservation of energy that for any change in the system

$$\left.\begin{matrix}\text{the increase in}\\ \text{internal energy}\end{matrix}\right\} = \text{the heat added} + \left\{\begin{matrix}\text{the work done}\\ \text{on the system.}\end{matrix}\right.$$

This is in fact merely a generalization (applied to a system of bodies instead of to a body) of the statement of §291 that the heat added to a body = the increase in internal energy + external work done by the body. The essential idea of the *first law of thermodynamics* is that *energy may be changed in form but cannot be destroyed.*

342. Isothermal Processes.—Any process or change of condition in a system which takes place without change in temperature is an isothermal process. We must distinguish between an isothermal process and an isothermal curve for a substance. Suppose the substance is in the gaseous state, then we have seen (§281) that, at a given absolute temperature T, the possible pressures and volumes are given very approximately by $PV = RT$, the isothermal curves being rectangular hyperbolæ. A gas having the pressure and volume determined by this equation, at the given temperature,

would, if confined in a cylinder with movable and properly weighted piston, be in *equilibrium*, that is, the piston would not move. All the conditions determined by the equation $PV = RT$ or the corresponding isothermal equation for a substance not a gas, are *equilibrium* conditions, the pressure being the equilibrium pressure corresponding to the given volume and temperature. It is evident that, to make a substance change its condition (T constant), the confining pressure must be changed from the equilibrium pressure, increased if it is desired to compress the gas, diminished if the gas is to be allowed to expand. If the change in pressure is very slight, the change in condition is slow; if the pressure is kept continuously slightly different from the equilibrium pressure given by $PV = RT$ = constant, the gas will pass through a series of conditions, in this case isothermally. The volumes of the gas, and the corresponding pressures exerted by the piston upon it, plotted on the PV diagram (Fig. 215), give the dotted curve just above and below the isothermal curve for the same temperature, and by making the process *slow* enough the dotted curve representing it may be made to approach as near as we wish to the equilibrium curve. We have seen that the work done upon the gas during isothermal compression is equal to the area under the isothermal curve between the extreme ordinates, and from the law of conservation of energy it follows that, neglecting the *change in internal energy* with volume, which we have seen (§292) is small, the equivalent of the work done upon the gas must be taken away as heat in order to keep the temperature constant. In a similar way, during isothermal expansion heat must be added.

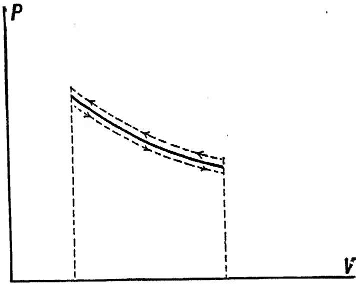


FIG. 215.—Isothermal equilibrium curve, and curves of reversible isothermal process.

343. Adiabatic Processes.—Any process carried out in such a way that no heat is allowed to enter or leave the system during the change, is called an adiabatic process. The same distinction as before exists between a process and a curve, an adiabatic curve determining a series of equilibrium conditions.

Through every point on the PV diagram one adiabatic and one isothermal curve will pass, the adiabatic being everywhere steeper

than the isothermal, because, since no heat is added, the temperature of the gas will fall as it expands and does work. Conversely, a substance has its temperature raised by adiabatic compression, the heat equivalent of the work done remaining in the substance, the work done being always represented by the area under the adiabatic curve between the extreme ordinates. The difference between isothermal and adiabatic compression may easily be illustrated by the use of a good bicycle pump, a slow compression being almost isothermal, the heat passing off as it is generated, through the metal walls of the cylinder, while quick compression warms the gas and cylinder considerably, as will be evident to the touch. Since it is impossible ever entirely to eliminate loss of heat by conduction, convection, or radiation, *quick* changes of volume will in general be more nearly adiabatic than slow. For this reason compressions and rarefactions in sound waves of low frequencies tend to be adiabatic; but at high frequencies the nearness of condensation and rarefaction and conduction between them neutralize the effect.

The equation of an adiabatic curve of a perfect gas, in the PV diagram, is

$$PV^{\kappa} = \text{constant}$$

κ being the ratio of the two specific heats, s_P/s_V. The same equation, having the same meaning, also holds very approximately for *real gases* which closely follow Boyle's law; and even for CO_2, which departs very considerably from Boyle's law, the adiabatic curve is given by the same *form* of equation, though κ is not in this case the ratio s_P/s_V.

344. The Equation of an Adiabatic.—In order to derive this equation we must discuss quantitatively the adiabatic and isothermal processes, and the exact relations of the quantities of heat and work which are involved.

Consider 1 gram of a gas which obeys Boyle's law at least approximately, confined in a cylinder with a movable piston.

Let V = the initial volume of the gas.
P = its pressure.
T = its temperature.

On the PV diagram, Fig. 216, this condition will be represented by the point a, recalling that, according to §272, each point represents a definite physical state. If the piston be so moved as to compress the gas by an amount $ab = \Delta V$, no heat being allowed to enter or leave the gas during this compression, then this process must, by definition, follow the adiabatic curve to the point d, both the pressure and temperature being increased, the final temperature being $T + \Delta T$. The condition d can also be reached by a combination of two

steps, a to b and b to d. First compress the gas at constant pressure along ab, during which process we must remove not only the heat equivalent to the work of compression (represented by area $abef$) but also additional heat in order to cool the gas through the interval $\Delta'T$, the total amount being $s_P\Delta'T$; then, keeping the gas at constant volume, add to it an amount of heat equal to $s_V(\Delta T + \Delta'T)$, thus bringing it to the condition d. Now since the points a and d represent definite physical states with definite amounts of energy the net energy expenditure in taking the gas from state a to state d *must be the same,* whether we go along ad or by $abcd$. In the first case the energy expenditure is the external work represented by the area $adbefa$ and in the second the work

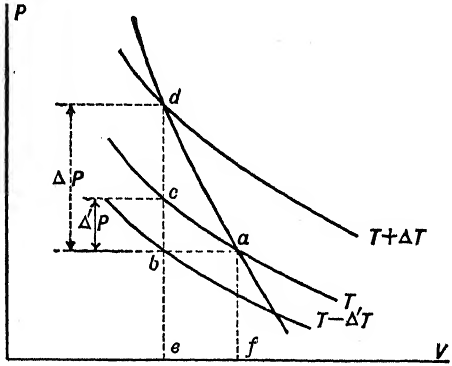


FIG. 216.—A small adiabatic change analyzed into the equivalent "volume constant" and "pressure constant" changes.

$abefa - s_p\Delta'T + s_v(\Delta T + \Delta'T)$. But the difference $adba$ between these two areas representing external work becomes relatively negligible as ΔV becomes very small, and accordingly it follows that

$$s_P\Delta'T = s_V(\Delta T + \Delta'T) \tag{1}$$

But (§264) the change in pressure of a gas confined at constant volume is very closely proportional to the change in temperature, hence

$$\frac{\Delta P}{\Delta'P} = \frac{\Delta T + \Delta'T}{\Delta'T}$$

and substituting from (1)

$$\frac{\Delta P}{\Delta'P} = \frac{s_P}{s_V} = \kappa \tag{2}$$

where κ is merely a symbol for the ratio $\frac{s_P}{s_V}$.

During this constant volume change the gas passes through a condition represented by c on the isothermal through a, and from the isothermal law (§223) it follows that

$$(P + \Delta'P)(V + \Delta V) = PV$$

(This is a general expression in which $\Delta'P$ and ΔV contain their own signs, since if $\Delta'P$ is positive ΔV must be negative, and vice versa.)

Hence
$$V\Delta'P + P\Delta V = 0 \tag{3}$$

Substituting for $\Delta'P$ from (2)

$$V\Delta P + \kappa P\Delta V = 0$$

or for infinitesimal changes we have

$$\frac{dP}{P} + \kappa\frac{dV}{V} = 0 \tag{4}$$

and, integrating,
$$PV^k = \text{constant}$$

345. Adiabatic Elasticity of a Gas.—The modulus of volume elasticity of a gas has been defined in §169 as

$$E = -\frac{dP}{dV}V$$

From equation (4) of §344 we see at once that for an adiabatic change of volume

$$E_{ad} = -\left(\frac{dP}{dV}\right)_{ad}V = \kappa P$$

346. Cyclic Operations. Reversible Processes.—A cyclic operation, or cycle, is a process or a series of processes so arranged that the system undergoing these changes is finally brought back to its initial condition. On the PV diagram any closed curve would evidently represent a cycle. Any such cycle may be divided into an expansion and a contraction (Fig. 217), and the area under the curve ABC represents the work done *by* the substances in the expansion ABC, while the area under the curve CDA represents the work done *on* the substance, during the compression CDA. The net work, in this case done *by* the substance, is evidently the area enclosed by the curve $ABCD$. If the cycle were described in the opposite sense, $ADCB$, the same amount of net work would be done *on* the substance. These conclusions are true whatever the form of the curve may be.

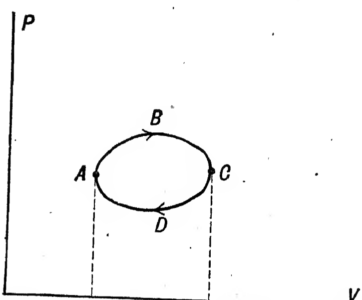


FIG. 217.—Curve representing a cyclic operation.

Any process is defined as reversible if it can be made to take place in the opposite sense by an infinitesimal change in the conditions, or, what is the same thing, if the curve representing the *process* (§342) lies infinitesimally near an equilibrium curve. For example, to make an isothermal process reversible, the pressure during expansion must always be infinitesimally near but less than, and during compression, infinitesimally near but greater than, the equilibrium pressure given by $PV = RT$, and the flow of heat must take place under an infinitesimal temperature gradient, that is *to* a body whose temperature is dT lower than that of the gas, or *from* a body whose temperature is dT greater than that of the gas. Under these conditions infinitesimal changes in P and T will cause the process to be described in the opposite direction. A cycle will be reversible if it is entirely made up of reversible processes.

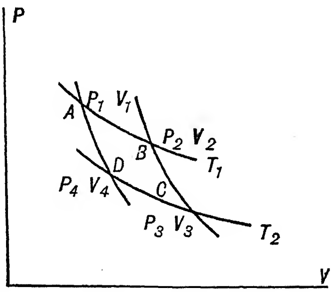


FIG. 218.—Carnot cycle.

347. The Carnot Cycle.—Carnot's Cycle, Fig. 218, is made up of two isothermal and two adiabatic processes, so chosen that the

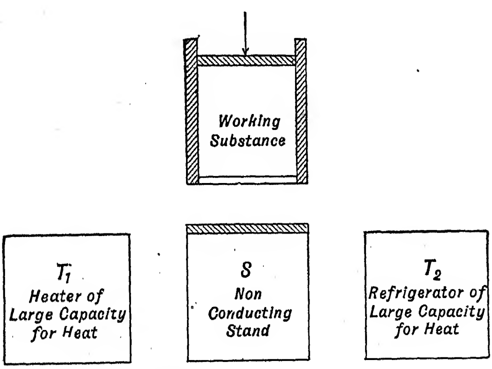


FIG. 219.—Carnot "engine," a device for using the Carnot cycle.

initial and final states are the same. Given a material, called the "working substance," conveniently (though not necessarily) a gas, inclosed in a cylinder with non-conducting walls and piston and a good conducting bottom (Fig. 219), together with a body (T_1) of very large heat capacity, at temperature T_1, a non-conducting stand

S, and a second body (T_2) of large heat capacity at temperature T_2, the Carnot cycle may be carried out as follows:

1. The working substance being initially in the condition P_1, V_1, T_1, (A, Fig. 218) place the cylinder on (T) and allow the gas to expand slowly to the condition P_2, V_2, T_1, absorbing heat by conduction from (T_1) during the process. If performed slowly, the process will be *isothermal* and *reversible*.

2. Place the cylinder on the insulated stand S, and allow the working substance to expand adiabatically (and reversibly) to the condition P_3, V_3, T_2.

3. Place the cylinder on the refrigerator (T_2) and compress isothermally to the condition P_4, V_4, T_2, heat being given off during the process to body 2. This will also be reversible if the compression is slow.

4. Place the cylinder again on the insulating stand and compress adiabatically to the initial condition.

According to §346 the net work, W, done by the gas when the cycle is described in this sense is represented by the area $ABCD$.

Let H_1 = heat taken in at temperature T_1 in mechanical units,

H_2 = heat given out at temperature T_2 in mechanical units.

Then according to the first law

$$W = H_1 - H_2.$$

If the cycle were carried out in the reverse order, then

H'_1 = heat given out at temperature T_1,

H'_2 = heat taken in at temperature T_2,

W' = work done on the gas during one cycle,

and again

$$W' = H'_1 - H'_2$$

and $W = W'$.

In order that the cycle may be reversed it is necessary, as we have seen, that the heat flow should take place with infinitesimal temperature gradient and the pressures always be infinitely near equilibrium pressures. The first condition can be satisfied as nearly as we wish by making the isothermal transformations slow enough, and the second condition by properly altering the force on the piston. The process which we have described, which enables us by means of a reversible Carnot cycle to get mechanical work from heat, is that of an *ideal heat engine*. We are not concerned at present with the mechanical construction of such an engine; the essential characteristic is the reversible Carnot cycle,

and one ideal engine differs from another only in the material used as working substance, and in the temperatures and pressures at which it works. An engine working between the temperatures T_1 and T_2 in which the flow of heat to and from the working substance took place under finite temperature gradients, or in which the forces on the piston were not properly adjusted during expansion and compression, or both, would be irreversible.

348. Efficiency of an Engine.—The efficiency of an engine is the ratio of the mechanical work obtained to the heat taken in by the working substance, during one cycle. For the Carnot cycle this is

$$e = \frac{W}{H_1} = \frac{H_1 - H_2}{H_1}$$

The efficiency gives the fraction of the heat taken in which is transformed into mechanical work.

349. The Second Law of Thermodynamics.—The second general principle of thermodynamics was first formulated independently by Clausius (1850) and Kelvin (1851) in equivalent but different forms, as follows:

It is impossible for a self-acting machine to convey heat continuously from one body to another at a higher temperature (*Clausius*).

It is impossible by means of any continuous inanimate agency to derive mechanical work from any portion of matter by cooling it below the lowest temperatures of its surroundings (*Kelvin*).

These are equivalent *axioms* or *assumptions*, which it is impossible to prove directly, but which are to be accepted as a basis of reasoning until some deduction from them is found to contradict fact. No such contradiction has ever been found. The second law recognizes and expresses a certain *natural tendency* of events, for example the tendency of heat to flow *down* a temperature gradient and of a compressed gas to *expand*. Stated in another way, it expresses the easily accepted generalization that *natural* processes, that is, processes which take place without assistance or control, are in general *irreversible*, as we have used the term.

The importance of the Second Law of Thermodynamics lies in the fact that it indicates clearly the only conditions under which heat can be converted into mechanical work. As the Kelvin statement makes clear, heat cannot be converted into work if everything is at the same temperature, for the conversion of heat into work

would necessarily mean the abstraction of heat from some part of the system and hence the lowering of its temperature below that of the surroundings. Accordingly *two different temperatures* are necessary, a higher and a lower—the heater and refrigerator temperatures of the Carnot cycle, the intake and exhaust temperatures of an engine. Claude's recent experiments in utilizing the heat energy of the warm surface waters in the southern ocean furnish a good example. A huge pipe was sunk to the ocean bottom to render available water of a lower temperature, and a special form of engine was arranged to work between these two temperatures. As we shall see later (§351), the efficiency of any heat engine is directly dependent on the *difference* between these two temperatures. The "heat death" of the universe, sometimes discussed, is the possibility—or probability—that all matter will eventually come to the same temperature and further conversion of heat into mechanical work will accordingly be impossible.

350. Carnot's Theorem.—We can now prove an extremely important theorem, which was first stated in 1824 by Carnot as follows:

The efficiency of all reversible engines taking in and giving out heat at the same two temperatures is the same, and no irreversible engine working between the same two temperatures can have a greater efficiency than this.

Carnot's proof of this theorem was incorrect, being based on the caloric theory of heat. As given by Clausius and Kelvin it is a necessary consequence of the second law. First consider any two reversible ideal engines, E and E', working between the temperatures T_1 and T_2, and let E' run *backward*. Let H_1 and H_2 be, as before, the heat taken in and given out by the forward-running engine, and H'_1 and H'_2 the heat given out and taken in by the engine running *backward*. Also let the engines be so connected mechanically, and of such a size or speed that the work done by the forward-running engine just suffices to operate the backward-running engine. Finally, let us assume for the moment that the efficiency of the forward-running engine is greater than the efficiency of the backward-running one. Then

$$e = \frac{H_1 - H_2}{H_1} > \frac{H_1' - H_2'}{H_1'} = e' \tag{1}$$

from the inequality of efficiencies, and

$$W = H_1 - H_2 = H_1' - H_2' = W' \tag{2}$$

from the equality of the work done by and on the engines, respectively. Hence from (1) and (2)

$$\frac{1}{H_1} > \frac{1}{H_1'}$$

or
$$H_1 < H_1'$$
and from (2)
$$H_2 < H_2'$$

Hence, the net result is that an amount of heat equal to

$$H'_2 - H_2 = H'_1 - H_1$$

is transferred from the body at the lower temperature T_2 to the body at the higher temperature T_1, *without the necessity of doing any work*. This violates the Clausius statement of the second law, hence we conclude that e cannot be greater than e'. If we run engine E' forward and E backward, we can prove by exactly similar reasoning that e' cannot be greater than e, hence it follows that $e = e'$, which proves the first part of the theorem.

If engine E is an *irreversible* engine, then we can prove exactly as above that e_{ir} cannot be greater than e', but since E cannot be reversed, we *cannot* prove that e' cannot be greater than e_{ir}. Hence all we can say is that e_{ir} is equal to or less than e_{rev}

$$e_{ir} \leqq e_{rev}$$

which proves the second part of the theorem.

351. Thermodynamic Scale of Temperature.—Since the efficiency of a reversible engine is independent of the working substance and the pressures used, it follows that the efficiency can depend only on the two temperatures between which the engine works. If $\dfrac{H_1 - H_2}{H_1}$ depends only on the temperatures T_1 and T_2, then $\dfrac{H_2}{H_1}$ also depends only on the temperatures. This fact led Lord Kelvin to suggest a new scale of temperature, which, since it depends on Carnot's theorem and is independent of the properties of any particular substance, is called the *absolute thermodynamic scale of temperature*. According to this scale, *any two temperatures are to each other as the heat taken in and given out by a reversible engine describing a Carnot cycle between these two temperatures*. That is, if we call θ_1 and θ_2 the thermodynamic measure of two temperatures,

$\frac{\theta_2}{\theta_1} = \frac{H_2}{H_1}$. We still have to determine the size of the degree, which is done, as in the case of the hydrogen centigrade scale by assuming 100° between the freezing-point and the boiling-point of water. This amounts to dividing the area (Fig. 220) between the 0° and 100° isothermals and any two adiabatic curves into one hundred equal parts, as indicated in the figure.

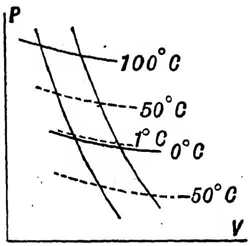


FIG. 220.—Thermodynamic temperature scale; difference in temperature proportional to work done (area).

We have then the following equations with which to determine any temperature θ on the absolute thermodynamic centigrade scale

by definition, $\frac{\theta}{\theta_{100}} = \frac{H}{H_{100}}$

and also $\frac{\theta}{\theta_0} = \frac{H}{H_0}$

and from the "centigrade" assumption,

$$\theta_{100} - \theta_0 = 100$$

from which, eliminating θ_{100} and θ_0, we get

$$\theta = \frac{H}{\frac{H_{100} - H_0}{100}}$$

If there were a reversible ideal engine actually available, the method of determining thermodynamic temperatures would be first to work the engine between the steam-point and the ice-point and determine the amount of work it could do, then to work it between the ice-point and a source of available temperature, such as a large tank of water, and adjust the temperature of the tank until the work done was $\frac{1}{100}$ of the amount done from 100° to 0°C. The tank would then be 1° above the ice-point on the *centigrade thermodynamic scale.* A similar process would determine other temperatures. Considering further the operation of an engine as represented by Fig. 220, it is evident from what has been said that the difference between any two thermodynamic temperatures θ_1 and θ_2 is proportional to the difference $H_1 - H_2$, which is the work done by the engine working between these temperatures and is represented by the area between the correspond-

ing isothermals and the two adiabatics. Hence 1° corresponds to $\frac{1}{100}$ of the area between the adiabatics and the isothermals of the ice and steam-points, the isothermal for 50° above the ice-point must divide this area in half, and an equal half area must be included between the isothermal of the ice-point and that for 50° below.

It also follows at once that the efficiency of a reversible engine may be expressed in terms of thermodynamic temperature as

$$e = \frac{H_1 - H_2}{H_1} = \frac{\theta_1 - \theta_2}{\theta_1}$$

Experiments have shown that, within the range of temperatures used by engines, θ and T are practically the same, so that we have, as a very close approximation,

$$e = \frac{T_1 - T_2}{T_1}$$

This is an equation of very great importance in its applications, since it indicates how to improve the efficiency of heat engines. It is accomplished primarily by having the largest possible difference between the intake and exhaust temperatures. In steam engines this means higher pressures and superheated steam, also the use of a condenser to cool the exhaust. By the use of mercury instead of water a larger $T_1 - T_2$ is obtained, but this is partly offset by the larger T_1. However there is still a net gain. The same may be said of internal combustion engines.

352. Comparison of Thermodynamic and Hydrogen Scale.—The *thermodynamic or Kelvin temperature scale* is entirely distinct from the hydrogen scale, and if it is to be adopted as the standard we must have either a practicable way of measuring in terms of it, or a way of comparing other scales with it.

It can be proved theoretically that the temperature indicated by a gas thermometer operating with a perfect gas would agree exactly with the thermodynamic temperature as defined above, using the perfect gas as the working substance. Unfortunately there is no perfect gas available for use in a thermometer; but as we have already pointed out, the properties of real gases approach those of a perfect gas as their densities approach zero. Accordingly, if a given gas, for example hydrogen, is used in a thermometer at several densities and the corresponding temperature scales are compared, the scale obtained by extrapolating from these to a condition of zero density will agree with the thermodynamic scale. Moreover, real gases differ from perfect gases in several important ways, namely:

1. The law $PV = RT$ is obeyed *exactly* by a perfect gas, but only *approximately* by real gases.

2. The internal work of free expansion is *zero* for a perfect gas, but is *not zero* for real gases.

3. The specific heat at constant pressure is *constant* for a perfect gas, but is *not constant* for real gases.

Hence, by measuring the pressure and volume of a real gas at various constant temperatures, by performing the porous plug experiment (§296) with it, and by measuring its specific heat under various conditions, it is possible to determine its "degree of imperfection," so to speak, and thence the relation between its constant volume temperature scale and the thermodynamic scale. There is still lacking much information concerning the properties of real gases, especially concerning the internal work of free expansion, due to molecular forces. Nevertheless the reduction to the thermodynamic scale is known with considerable accuracy for both hydrogen and nitrogen, as given in Table 18.

TABLE 18

CORRECTION FOR CONSTANT VOLUME THERMOMETER SCALES

$P_o = 1000$ MM. HG.

Temperature (Centigrade).	Nitrogen.	Hydrogen.
− 250		+0.12
− 200	+0.5	+0.06
− 150	+0.2	+0.03
− 100	+0.06	+0.015
− 50	+0.03	+0.005
0	0.000	0.000
+ 25	−0.008	−0.001
+ 50	−0.010	−0.002
+ 100	0.000	0.000
+ 200	+0.02	+0.02
+ 500	+0.2	
+1000	+0.7	
+1200	+1.0	

It is evident that for moderate temperatures and approximate work the thermodynamic and hydrogen scales may be considered identical.

353. Entropy.—Returning now to the Carnot cycle we see that, as a result of the definition of the thermodynamic scale of temperature, we have

$$\frac{H_2}{\theta_2} = \frac{H_1}{\theta_1}$$

This means that the ratio of the heat taken in (or given out) to the temperature at which it is taken in (or given out) is the same for all isothermal changes between any two adiabatics. This fact suggested to Clausius that the quantity $\frac{H}{\theta}$ is the change in a certain property of the working substance, a property which remains constant during any (reversible) adiabatic process but changes when the substance passes from one adiabatic to another. This property Clausius named "entropy," and it is exceedingly important.

In order to obtain a definite numerical measure for the entropy of a body in every physical condition, we must select some condition, represented by a point on the PV diagram, as an arbitrary zero of entropy, just as we select the sea level as the zero from which to measure heights and depths. Suppose P (Fig. 221) is the adopted zero, then the entropy of any other state P' is obtained by measuring the heat taken in (or given out) in passing from P to P' by a *reversible* path. The simplest path is by the adiabatic PN and the isothermal θ. If H is the heat taken in in passing from N to P', then the entropy of P' with respect to P, which we shall represent by $S(P, P')$, would be equal to $\frac{H}{\theta}$. If P' were reached by another reversible path involving portions of several adiabatics and isothermals, and quantities of heat $H_1, H_2, H_3, \ldots$ were taken in (or given out) at the temperatures $\theta_1, \theta_2, \theta_3, \ldots$ then $S(P, P') = \frac{H_1}{\theta_1} + \frac{H_2}{\theta_2} + \frac{H_3}{\theta_3} + \cdots = \Sigma(H/\theta)$. If any of the quantities of heat $H_1, H_2, \ldots$ were given out by the body, they should be taken with the minus sign in the summation. It is evident that, defined in this way, every state has a definite entropy.

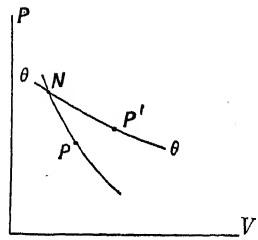


FIG. 221.—Arbitrary zero of entropy, P; Entropy of P' determined by adiabatic-isothermal change from P.

354. Entropy and Reversible Cycles in General.—We have seen that in passing around a Carnot cycle the entropy of the working substance was not changed. This result may be extended to include any reversible cycle, represented by the closed curve in Fig. 222. By drawing a series of adiabatics across this and connecting these around the edge by a series of isothermal steps as shown, we see that the given cycle may be broken up into a series of Carnot cycles, the sum of whose areas will approach the area of the given cycle as a limit, as their number is increased. Furthermore, the heat taken in along the isothermal steps AA', BB', CC', DD', etc., is equal in the limit to the heat taken in moving along the curve $A - D$, for the difference would be represented by the sum of the triangular areas, which is zero in the limit. Hence, since for each elementary Carnot cycle we have

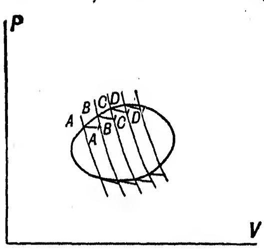


FIG. 222.—Analysis of any cycle into elementary Carnot cycles.

$$\frac{H_1}{\theta_1} = \frac{H_2}{\theta_2}$$

for the whole cycle
$$\Sigma(H_1/\theta_1) - \Sigma(H_2/\theta_2) = 0$$
and, if we give the negative sign to heat H_2 *leaving* the system, this becomes,
$$\Sigma(H/\theta) = 0$$
or, finally,
$$\int\frac{dH}{\theta} = 0$$
when the number of elementary cycles has become infinite. This shows us that all reversible paths between two conditions involve the same change in entropy, or that $\frac{\Delta H}{\theta}$ is a *perfect differential.*

355. Increase of Entropy.—If an amount of heat H flows from one body at a temperature θ_1 to another at a lower temperature θ_2, the entropy of the hot body is decreased by an amount $\frac{H}{\theta_1}$ and that of the cooler body is increased by $\frac{H}{\theta_2}$. Evidently in all cases of conduction, $dS = H\left(\frac{1}{\theta_2} - \frac{1}{\theta_1}\right)$, is *positive*, or the entropy of the two bodies is increased.

It can also be proved that other "natural" processes such as free or unbalanced expansion, the diffusion of gases into each other, dilution of a solution, and the production of heat from mechanical energy by friction, all involve an increase in entropy. These processes are also all irreversible, and they all

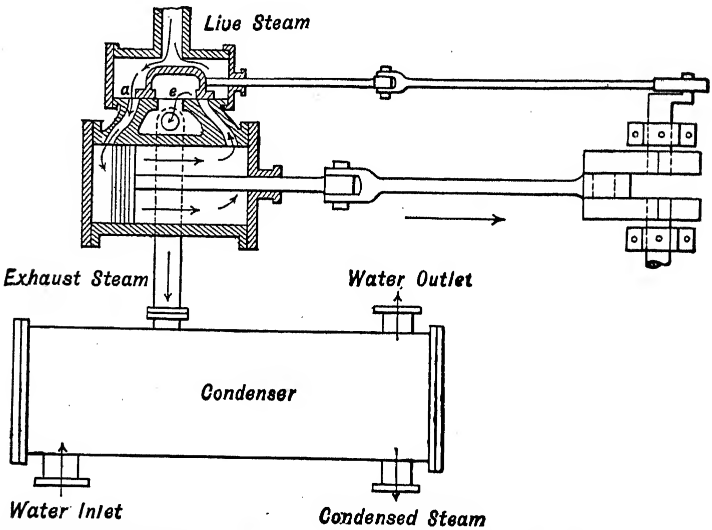


FIG. 223.—Reciprocating steam engine and condenser in which steam is condensed on water-cooled pipes.

tend to a more uniform condition as regards temperature, pressure and the velocities of bodies and of molecules. Hence it is a reasonable extension of our ideas to say that *all* natural processes are irreversible and lead to an increase in entropy, and to associate the increase of entropy with increase in the *uniformity of physical conditions*. All natural changes seem to be tending to a condition of *maximum uniformity*. This additional hypothesis, that natural processes always lead to an increase in entropy, is the basis for the discussion of problems of chemical and physical equilibrium, such as the equilibrium of a liquid with its vapor, of a solid with its liquid, or of different chemical compounds with each other.

356. Reciprocating Steam Engines.—The ordinary reciprocating steam engine, one type of which is shown in Fig. 223, is the

most common machine used to convert heat energy into mechanical work. In these engines water is the *working substance* (compare §347), the boiler is the source of heat at the higher temperature, and the cooling water of the *condenser* is the cooler body into which heat is discharged. Instead of moving the cylinder from one to the other as was before suggested, it is obviously easier to conduct the working substance from point to point. On account of mechanical difficulties no attempt is made to realize completely the Carnot cycle (§347), but the actual cycle through which the working substance passes is of the form shown in Fig. 224. The operations are as follows:

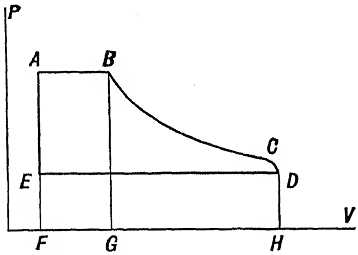


FIG. 224.—Ideal (Rankine) cycle for a reciprocating steam engine.

(1) Water is vaporized in the boiler at the temperature T_1, absorbing an amount of heat L_1 per unit mass (heat of vaporization).

(2) Steam passes at constant pressure P_1 from the boiler through the valve a (Fig. 223), into the cylinder as the piston begins its motion to the right. Thus the isothermal expansion at pressure P_1 due to vaporization is represented by the line AB, and this expansion does an amount of work represented by $ABGF$.

(3) The valve a closes, and the saturated steam expands from B to D. This expansion should be as nearly as possible adiabatic. The work done is represented by the area $BDHG$. At C the valve a opens to the exhaust e, the steam begins to escape to the condenser, and the pressure falls quickly from C to D.

(4) The piston reverses its motion at D, and the motion to the left is opposed by the constant pressure P_2, since during this time there is isothermal condensation of the steam in the condenser, at temperature T_2. The temperature T_2 is fixed by the cooling water which is available for the condenser. With a non-condensing engine T_2 is necessarily about 373° absolute (100°C.). An amount of heat L_2 per unit mass is given up to the cooling water during the process of condensation, and work to the amount $DEFH$ is done *on* the steam.

(5) The condensed steam is heated at constant volume (EA) and admitted to the boiler at A, thus completing the cycle. This requires an additional amount of heat H.

It is possible to arrange a mechanism so that the engine as it runs will automatically draw a curve whose ordinates are proportional to the pressure of steam in the cylinder, and whose abscissæ are proportional to the corresponding volume occupied by the steam in the cylinder. This curve is very similar to the one of Fig. 224, and is called an *indicator* diagram.

357. Efficiency of Engines.—The work, W, done per pound of steam is evidently represented by the area $ABDE$, while the total heat taken in is $L_1 + H$. The ratio $\frac{W}{(L_1 + H)J}$, where J is the mechanical equivalent of heat, is called the *thermal efficiency* of the engine. The thermal efficiency measures the perfection of the thermal processes which the engine uses. There is, of course, energy lost (converted into heat), by friction among the moving parts, so that the actual work, W', which the engine could do in running some machine, is always less than W. The ratio $\frac{W'}{W}$ is called the *mechanical efficiency* of the engine, and its value is a measure of the mechanical perfection of the engine. The product of the two efficiencies, namely, the ratio $\frac{W'}{(L_1 + H)J}$, evidently measures the efficiency of the engine in the conversion of heat into *usable* mechanical work, and this will be less than its thermal efficiency.

It is interesting to compare the thermal efficiency $\frac{W}{(L_1 + H)J}$ with the efficiency of an ideal Carnot engine working between the same temperature T_1 and T_2.

Since, according to §352, the constant volume hydrogen scale and absolute thermodynamic scale of temperature are practically identical, the expression for the efficiency of an ideal engine becomes $e = \frac{T_1 - T_2}{T_1}$, and this is the maximum efficiency which any real engine could possibly be expected to approach if it works with a boiler temperature T_1 and a condenser temperature T_2. For example, with a boiler at 177°C. and a condenser at 77°C., $e = \frac{100}{450} = 22$ per cent., that is to say, the ideal engine could convert less than *one-quarter* of the heat used into mechanical work. Table 19 gives the actual thermal efficiency and the corresponding ideal efficiency for the best engines of several types.

Besides engine efficiencies, the efficiency of boilers, namely, the ratio $\frac{\text{heat given to water}}{\text{heat obtained from fuel}}$ (in a given time) is of course of equal importance in the problem of obtaining mechanical work from fuel. The average efficiency of boilers is perhaps 60 per cent., the maximum 90 per cent., so that, combining the best boiler with the best engine, the maximum efficiency actually attained is about 21 per cent. Turbines exceed this somewhat. See Table 19.

From §298 the heat of combustion of soft coal is 2.9×10^{11} ergs per gram, or 12,500 B. T. U. per pound, while (§61) one horse-power for one hour equals 2.68×10^{13} ergs, or 1.98×10^{6}

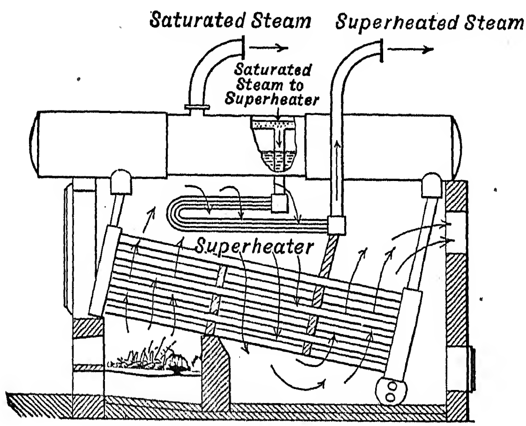


FIG. 225.—Boiler and superheater.

foot lbs. Since 1 B. T. U. equals 778 foot lbs., it follows that the combustion of 1 lb. of coal liberates energy sufficient to provide 4.8 H. P. for one hour, whereas the best boiler-engine combination so far built obtains perhaps 1.6 H. P. hour per lb. of coal.

358. The Defects of Real Engines.—In the discussion of §357 we have neglected several points of importance. For example, the expansion BC can never be strictly adiabatic because the cylinder and piston must be of conducting material. This leads to the condensation of steam in the cylinder. If it is attempted to raise the temperature T_1 so as to increase the efficiency, the cylinder condensation is increased. By using several cylinders (compound, triple and quadruple), allowing part of the expansion to occur in each, the temperature changes in each cylinder, and hence the condensation losses, are reduced and it is possible to use higher initial temperatures. Further reduction of condensation loss and increase of the initial temperature, without increase

in the initial pressure, is accomplished by *superheating* the steam, by passing it, at constant pressure, through coils of pipe in the hot flue gases, as shown in Fig. 225. It is then no longer saturated when it enters the cylinder, and the cycle would be represented by different lines on the PV diagram.

359. Steam Turbines.—The turbine is another type of machine, of more recent development, for obtaining mechanical work from the heat energy of

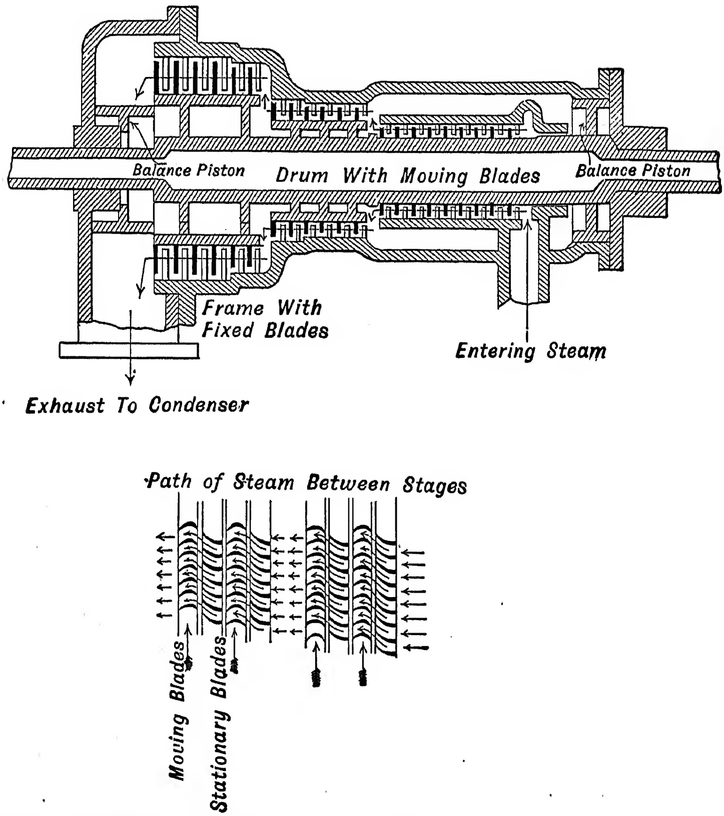


FIG. 226.—Steam turbine, pressure type; general arrangement and detail showing flow of steam past blades.

steam, the essential features being a rotating shaft with properly arranged blades and fixed nozzles or blades for directing the flow of the steam, which is initially at a high pressure. Turbines may be divided into two general classes. In the first class, called the "velocity" type, steam is allowed to expand at once to the final pressure, in a properly shaped nozzle, so that the jets acquire a high velocity. These jets impinge on the movable blades and cause them to rotate, much as the jets of water impinge on the blades in certain types of water wheels (§204). By using several sets of movable blades, with fixed

passages between for reversing the direction of the steam jet, the drop in velocity is rendered more gradual, and the speed of the turbine shaft need not be so great. In the second class of turbines, called the "pressure" type, shown in Fig. 226, the steam expands gradually through a great many sets of movable and fixed blades, exerting a pressure on each set which causes the movable blades to rotate. Steam turbines have certain mechanical advantages over reciprocating engines, namely, uniform and high angular velocity, freedom from

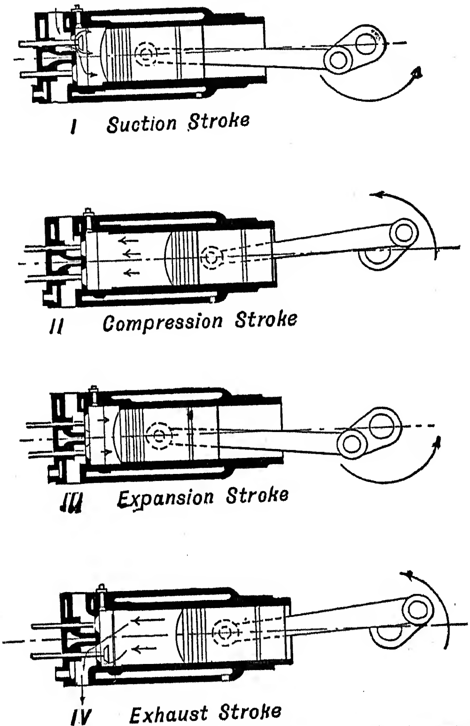


FIG. 227.—Four-cycle internal combustion engine, showing the four stages of one cycle.

vibration (hence their desirability for use in steamships), and also economy of space. Turbines of over 110,000 kw. capacity have been built, taking steam at a pressure of 1200 lbs. per sq. in.

360. Internal Combustion Engines.—In these engines the function of the boiler and expanding cylinder are combined, the combustion taking place in the cylinder itself. The way in which this is carried out in the "four cycle" type of engine can best be understood by describing the various stages shown in Fig. 227. In I the inlet valve is open during the entire stroke to the right, admitting a cylinder full of a proper explosive mixture of a combustible (coal gas, gasoline or alcohol vapor) and air. In II the valve is closed and the return stroke is taking place; this compresses the mixture into the clearance space at the end of

the cylinder, which is called the explosion chamber. At the end of the compression stroke the mixture is exploded, usually by an electric spark. The high pressure resulting from the combustion acts upon the piston during the stroke III to the right, while at the end of this stroke the exhaust valve e opens, and during IV the products of combustion are expelled preparatory to beginning over again as in I. Engines using the series of operations just described are called "four-cycle" engines, because four strokes are necessary to complete the series. There are other types of engines, notably the "two-cycle," requiring only two strokes to complete the series of operations, and the Diesel, in which air alone is compressed and the fuel is injected into it, the result being quiet combustion, instead of an explosion as in the four-cycle type. The thermal efficiency of the best four-cycle engines is about 30 per cent., of the Diesel type about 40 per cent. Aside from high efficiency, internal combustion engines have the further advantages of compactness, ease of handling and quickness of starting. The formation of the proper mixture of fuel and air is the most troublesome operation in the running of an internal combustion engine, this being usually accomplished in separate attachments called carburetors. The largest engines of the internal combustion type so far built are of 25,000 H. P.

TABLE 19

EFFICIENCY OF ENGINES

(Efficiencies marked * must be multiplied by a boiler efficiency of 85–90 % to give overall efficiency)

Engine	Temperature t_1	Temperature t_2	Efficiency, per cent	Efficiency of Carnot cycle, per cent
Steam				
Willan's engine (non-condensing)..	164°C.	101.5°	10.4*	14.5
Levitt pumping engine (compound)	181.6°	37.7°	19 *	31.7
Levitt pumping engine (triple expansion)...................	191.9°	46.7°	20.8*	31.8
Nordberg engine (quadruple expansion).........................	206.3°	43.1°	25.5*	34.0
Large turbines, high pressure steam	400°	30°	35*	55.0
Mercury turbine—steam turbine combination..................	......	......	35	
Internal combustion				
Ordinary automobile engine.......	......	......	22	
Large Diesel oil engine...........	1650°	600°	40	54.5

361. Mechanical Refrigeration.—The refrigerating machine is an interesting and important application of the principles of the Second Law of Thermodynamics. In this case the cycle is reversed and the machine acts as a "heat pump," taking in the heat H_2 at the lower temperature and giving out H_1 at

the higher. The difference W is the energy which would have to be expended in running a theoretically perfect machine of this sort working in a Carnot cycle (actually the compressor works in a Rankine cycle somewhat like that of the steam engine). Since

$$\frac{H_1 - H_2}{H_2} = \frac{T_1 - T_2}{T_2} \text{ we have } W = H_2\frac{T_1 - T_2}{T_2}.$$

If the refrigerating coil is at $-15°$C. (258°K.) and the condenser at $+40°$C. (313°K.), we have $H_2/W = \frac{258}{55} = 4.70$. A power expenditure of 50 watts or 43,000 calories per hour would in this case abstract 202,000 calories per hour from the refrigerating coil. As a matter of fact, however, the efficiency of an actual machine is of the order of only 25% of this theoretical efficiency, so 50 watts might take 50,000 cal./hr. from the refrigerator.

The working substance is ammonia in large plants (§321) and some material such as sulphur dioxide in household refrigerators. This is compressed and condensed to a liquid and cooled (Fig. 204). The liquid then goes through an expansion valve into the refrigerating coil where it vaporizes, absorbing as it does so the latent heat of vaporization. It is then drawn into the compressor again and the cycle is repeated. The degree of cooling is determined by the regulation of the expansion valve, which is usually controlled by a thermostat.

In addition to refrigerators using electrically driven compressors there are also gas or kerosene operated cooling devices, using ammonia as a refrigerant. In one type of these the burners are lighted for an hour or two each day and ammonia gas driven from a mixture of ammonia and water in a small boiler or generator into water-cooled coils, where it condenses. When the flame goes out and the pressure is reduced the ammonia re-evaporates in the refrigerating coil, absorbing heat as it does so, and eventually returns to the generator. In another type of gas refrigerator the action is a continuous one; the principle is the same but the cycle is more complicated.

362. Air Conditioning.—By air conditioning is meant the bringing about of such conditions of temperature, humidity, and circulation (sometimes also filtration and deodorizing) of the air in a room as will afford maximum comfort. While the heating is usually produced by the combustion of coal, etc., the use of the reversed cycle or "heat pump" has interesting possibilities. The principle is exactly that of the mechanical refrigerator, H_2 calories being abstracted from an absorber coil outside, corresponding to the refrigerator coil, and H_1 given to a radiator inside. If these temperatures are, say 10°C., and 40°C., we see from the considerations of the preceding article that $H_1 = \frac{313}{30}W$, or the expenditure of 1 kw.-hr. of energy W would give to the radiator over 10 kw.-hr. or nearly 9×10^6 calories in a theoretically perfect machine. The actual heat realized would, of course, be much less than this, but the method is practical for small temperature differences. There are a few installations of this sort in mild climates, arranged to cool the building in summer and heat in winter, using the radiator as an absorber and vice versa.

Cooling in air conditioning is generally brought about by a special refrigerating machine. In some cases the air is cooled well below the final temperature, dehumidified by condensation of the moisture, and then heated to the desired temperature. In this way any desired relative humidity may be

obtained. Thus if air is to be delivered at 22°C. (sat. vap. pressure 19.8 mm. Hg.) and 70% relative humidity, it should first be cooled to about 16°C. (v.p. 13.6 mm.) and then warmed to 22°.

PROBLEMS

Temperature. 1. Find the value of the following temperatures on the Centigrade scale: The temperature of the human body (98°F.); normal temperature of a living room 68°F.; a cold day in winter (20°F. below zero). *Ans.* 36.6°C.; 20°C., −28.9°C.

2. Find the value of the following Centigrade temperatures on the Fahrenheit scale: Absolute zero; temperature of "dry ice" (−78.5°C.); melting-point of gold; temperature of surface of sun. *Ans.* −459.8°F.; −109.3°F.; 1945°F.; 10832°F.

3. At what temperature do the Fahrenheit and Centigrade thermometers read the same? The Fahrenheit twice the Centigrade? *Ans.* −40°; 160°C.

Expansion. 4. A clock which has a pendulum made of brass keeps correct time at 20°C.; if the temperature falls to 0°C., how many seconds will it gain or lose per day? *Ans.* 16.3 sec. gain.

5. Steel rails in 40 ft. lengths are laid in winter at 0°C. How much space between consecutive rails must be allowed to permit expansion to a summer temperature of 45°C.? *Ans.* 0.23 in.

6. Steel street car rails, having their ends welded together, are laid in concrete so that it is impossible for them to move. Find the stress in the rails at −10°C., assuming that they are laid when the temperature is 20°C. *Ans.* About 7.2×10^8 dynes/cm.^{2}

7. Compute the change in volume of a block of cast iron 3 in. × 4 in. × 10 in. if the temperature changes from 44°F. to 116°F. *Ans.* 0.15 cu. in.

8. A uniform cylinder is filled with hydrogen under atmospheric pressure. This piston stands at a height of 400 cm. at 20°C. If the pressure is kept constant, find the height of the piston at the following temperatures; 100°C.; 300°C.; −80°C.; −180°C. *Ans.* 510 cm.; 783 cm.; 263.4 cms.; 127 cm.

9. If in the preceding problem the same gas is compressed until the piston stands at a height of 200 cm. and the volume is then kept constant, compute the pressure for the temperatures given in problem 8. *Ans.* 193.5; 297.0; 100; 48.2 cm. Hg.

10. Beginning at 10 atmospheres pressure and 2 liters volume, 10 g. of air has its pressure so changed that the p, v curve is a 45° straight line. Discuss the temperature changes which occur. (Fig. 187.)

11. Given 10 liters of nitrogen at 30°C. and 120 atmospheres pressure, what would be its volume at 100°C. and 200 atmospheres pressure? *Ans.* 7.4 liters.

12. What would be the relative increase in size of an air bubble in passing from the bottom of a lake 20 m. deep where the temperature is 4°C. to the top where the temperature is 20°C.? *Ans.* $V_2 = V_1 \times 3.1$.

13. The pressure in a tire is 35 lb./sq. in. in excess of atmospheric pressure (14.5 lb./sq. in.) when the tire temp. is 8°C. What will the gauge read when the tire temp. is 45°C.? *Ans.* 41.5 lb./sq. in.

14. What would be the lift of a balloon of 84,000 cu. ft. capacity filled at 20°C. to the pressure of 20 cm. of water above atmospheric pressure, (a) with hydrogen, (b) with helium? (See p. 105.) Compute the difference in lift in the case of hydrogen when filled and used at 0°C.? at 30°C.?

15. A certain stratosphere balloon is partially filled with 400,000 cu. ft. of gas at 75 cm. barometer and 20°C. What will the volume be when the temp. is −40°C. and the pressure 10 cm. mercury? *Ans.* 2.38×10^6 cu. ft.

16. The pressure in a certain vessel is reduced to 10^{-6} mm. mercury at a temp. of 17°C. How many molecules/c.c. remain in the vessel? (Number of gas molecules/c.c. at 0°C. and 760 mm. is 2.70×10^{19}.)
Ans. 3.34×10^{10}.

17. An evacuated 1 liter flask develops a crack through which one trillion molecules of air enter per second. How long before the pressure in the flask will rise to 38 cm. mercury, *i.e.*, half of atmospheric pressure?
Ans. 428 years.

Calorimetry. **18.** A 100 gm. piece of silver at 90°C. is dropped in 160 grams of water contained in an iron calorimeter weighing 40 grams. Temperature of water initially 15°C. Compute the rise in temperature of the water. *Ans.* 2.46°C.

19. 50 grams of a substance at 100°C. is dropped into 100 grams of water at 4°C. If the water is contained in a copper calorimeter, mass 60 grams, and the temperature of the water changes to 10°C., compute the specific heat of the substance. *Ans.* 0.140.

20. A water heater will heat 50 liters of water per minute from 15°C. to 80°C.; if the efficiency is 25 per cent., how many calories must be generated in the heater to do this? *Ans.* 13×10^6 cal.

21. How many liters of gas will be required per minute in the preceding problem? Assume density of gas as 0.0009 gm./c.c. and heat of combustion as 8000 cal./gm. *Ans.* 1805 liters.

Mechanical Equivalent of Heat. **22.** In drilling a hole in a block of iron, power is supplied at the rate of .8 H. P. for 3 minutes. How much heat is produced? If ¾ of this heat goes to warm the iron whose mass is 700 grams, find its change in temperature.
Ans. 25.6×10^3 cal.; 230°C.

23. A loaded truck weighing 6 tons and traveling 40 mi./hr. is stopped with the brakes. How much heat is developed? (1 B. T. U. = 778 ft./lbs.)
Ans. 825 B. T. U.

24. How much is the temperature of the water going over Niagara Falls (160 ft. high) raised by the impact? Does this apply to the water which goes through the turbines? Explain. *Ans.* 0.21°F.

25. Determine the heat produced in stopping a fly-wheel of 112 lbs. mass and 2 ft. in radius, rotating at the rate of one turn per second, assuming the whole mass concentrated in the rim. *Ans.* 0.354 B. T. U.

26. If electrical energy is 4 cents per 1000 watt hours and gas 90 cents per 1000 cu. ft., what is the relative cost of gas and electric heating? Take

density of gas as 0.0009 gm./c.c. and heat of combustion as 8000 cal./gm.
Ans. Cost of elect. = 10.5 cost of gas.

27. The solar constant is 1.93 cal. per sq. cm. per. min. and approximately $\frac{1}{3}$ of this radiation is absorbed by the atmosphere. How much horsepower is received per square yard of the earth's surface with vertical sun?
Ans. 1 H. P.

28. A sterilizer is filled with steam at a gauge pressure of 25 lb./sq. in. The gauge reads 0 at atmospheric pressure. What is the approximate temperature of the sterilizer? *Ans.* 130°C.

Change of State.

29. How much would the air in a room 6 × 5 × 3 meters be warmed by the condensation alone of 1 kg. of steam in the radiator? What would it be if the room were air-tight? *Ans.* 19.4°C.; 27.2°C.

30. With what velocity must a lead bullet at 50°C. strike against an obstacle in order that the heat produced by the arrest of its motion, if all produced within the bullet, might be just sufficient to melt it? *Ans.* 335 m./sec.

31. A specimen of copper with a mass of 400 g. is heated to 220°C. and placed on a mass of ice. If 100 g. of ice is melted, what is the specific heat of copper? *Ans.* 0.091.

32. How much steam at 150°C. must be added to 1 kg. of ice at −10°C. to give nothing but water at 0°C.? *Ans.* 128 grams.

33. A 200 g. piece of metal at 20°C. was introduced into a steam calorimeter and the steam turned on. The piece of metal, which was suspended by a fine wire running outside the calorimeter to a delicate balance, immediately was observed to gain in weight due to the condensed steam. When it had reached the steam temperature of 100° the weight became steady at 204.68 g. What is the specific heat of the metal? *Ans.* 0.158.

34. 5 kg. of ice at −30°C. is in a 2 kg. copper calorimeter at the same temperature. 1200 g. of steam superheated to 180°C. is introduced into this calorimeter. What is the final temperature? Take specific heat of ice as 0.5, copper 0.094, steam 0.48; heat of fusion of ice 80 cal./g.; heat of vaporization of water at 100°C., 540 cal./g. *Ans.* 52.5°C.

35. What is the relative humidity of air at 30°C. if the dew point is found to be 10°C.? *Ans.* 29.0 per cent.

36. A room 8 × 6 × 3 m. is at 20°C. and relative humidity 20%. How much water must be evaporated to raise the humidity to 40%? *Ans.* 498 gm.

37. How much heat would be required to convert 1 gm. of water-substance from liquid at 0°C. to vapor at 150°C. under 1 atmosphere pressure?
Ans. 664 cal.

38. Compute the "external" part of the heat of vaporization of water at 100°C.
Ans. 40.2 cal.

39. Carry a mass of substance across the triple-point diagram as shown on p. 249, explaining just what happens at the different points. Do this for both constant pressure and constant temperature following the dotted lines.

40. If it is desired to heat CO_2 at constant volume in a closed tube and have the substance pass through the critical point, what proportion of liquid and vapor must there be at 20°C. initially?
Ans. About $4\frac{1}{2}$ parts vapor to 1 of liquid.

41. Some ether is poured into a bottle containing air at atmospheric pressure and the bottle quickly corked; upon shaking the bottle and removing the cork a "pop" is heard. Explain.

42. Thermometers are used for indicating the temperature of aeroplane engines, which depend on the change in vapor pressure of ethyl ether. They consist of a metal bulb, a capillary connecting tube and a pressure gauge. How much ether must be put in the bulb in order to be sure that the pressure developed depends on the temperature of the *bulb* and not on the temperature of the rest of the instrument?

Heat Conduction.

43. The walls of a certain refrigerator have an area of 15,000 cm.^{2}, and are made of cork 3 cm. thick. Find out how much ice may be expected to melt in one day if the outside temperature is 86°F. and the inside averages 50°F. *Ans.* 14 kg.

44. The top of a steam chest containing steam at atmospheric pressure consists of a slab of stone 61 cm. long, 50 cm. broad and 10 cm. thick. The top being covered with ice, it was found that 4.8 kg. were melted in 39 minutes. What is the conductivity of the stone?
Ans. 0.0054 c.g.s. units.

45. One end of a copper bar 4 sq. cm. in cross-section and 80 cm. long, is kept in steam under one atmosphere pressure and the other end in contact with melting ice. How many grams of ice will be melted in 10 min.? Neglect loss due to radiation. *Ans.* 34.9 grams.

46. How much anthracite coal must be burned to make up for the loss of heat due to conduction for one day through a glass window 3 mm. thick and having an area of 3 square meters, supposing the air in the room to be at temperature 25°C., and the outside air at −20°C.? What important point has been neglected? *Ans.* 263 lb.

Radiation.

47. A black radiator 2 square meters in area, is in a room whose walls are at temperature of 18°C. If the radiator is at 100°C. at what rate does the room gain heat? The constant s of Stefan's law, $E = sT^4$, is 5.7×10^{-12} watts per square centimeter.
Ans. 332 cal./sec.

48. If the temperature of a furnace is measured by allowing the heat radiated through a hole 1 square centimeter in area in the walls to warm 100 grams of water placed in front of the hole, what is the temperature if the water rises in temperature by 13°C. in 1 minute? Assume that all the heat radiated from the hole is absorbed by the water and also neglect the heat radiated back into the furnace by the water. *Ans.* 1726°C.

Thermo-dynamics.

49. How many degrees will dry air at 15°C. rise in temperature if compressed adiabatically to $\frac{1}{5}$ of its volume?
Ans. 260°C.

50. How much work would be done by air in expanding adiabatically from the point P = 760 mm. Hg., V = 800 c.c. to the point P = 400 mm. Hg.? Solve graphically. *Ans.* 337×10^6 ergs.

51. Assuming the efficiency of an automobile engine as 22%, how much gasoline would be consumed per hour by an engine delivering 20 K.W. (26.8 H.P.)? Take heat of combustion of gasoline as 11,500 cal./gm. and density as 0.74. 1 U. S. gallon = 3.78 liters. *Ans.* 6.8 kg; 2.4 gal.

52. What is the total force on the end of a boiler 3 ft. in diameter if the temperature of the water inside is 180°C.? *Ans.* 149,000 lbs.

53. Compute the theoretical Carnot efficiency of (*a*) a steam engine, boiler temp. 160°C., condenser 50°C.; (*b*) a gasoline engine, explosion temp. 1400°C., exhaust 400°C. *Ans.* 25.4 %; 65.8 %.

54. A certain engine burns 100 lbs. of soft coal per hour. How much work would the engine do if all this heat were converted into mechanical work? In reality the engine furnishes 20 H.P. What is the efficiency of boiler and engine combined? *Ans.* Approx. 10^9 ft. lbs. per hr.; 3.9 per cent.

55. If an engine working at the rate of 622.4 H.P. keeps a train at constant speed for 10 minutes, how much heat is produced, assuming that all the work done is converted into heat? *Ans.* 6.65×10^7 cal.

56. A certain hoisting engine runs at 800 R.P.M. and takes in 10,000 cal. of heat at each revolution. Boiler temp. 180°C., exhaust 102°C. If its efficiency is 52 % of that of a similar Carnot engine compute the power developed in K.W. and in H.P. (1 H.P. = 746 watts). *Ans.* 50 K.W., 67 H.P.

57. What must be the boiler efficiency in order that a Nordberg quadruple expansion engine should furnish 1 H.P. by burning 1 lb. of soft coal per hour? *Ans.* 88 per cent. for average soft coal.

58. A certain electric refrigerator has a heat leakage of 200,000 cal./hr. If the cooling element runs at −10°C. and the radiator at +40°C., and if the refrigerating plant has an efficiency of 40 % (of that of an ideal Carnot refrigerator) what power will be taken by the motor? *Ans.* 110 watts.

59. Can a kitchen be cooled by leaving the door of the electric refrigerator open? Explain.

60. A building, requiring 10^8 calories/hour, is to be heated by means of the reversed cycle (see Art. 362). If the outside absorber coil is at 5°C. and the inside radiator at 50°C., and if the plant has an efficiency of 30 % of that of a Carnot engine, what power expenditure will be required? *Ans.* 54 K.W.

ELECTRICITY AND MAGNETISM

By Reginald J. Stephenson, Ph.D.

University of Chicago

363. We are all acquainted, to some extent, with electrical ienomena, because of the extensive use of electricity in the world ound us. Street cars and electric motors show that electricity n be used to supply mechanical power. Electric heaters and mps illustrate how it can be applied to produce heat and light. adio is a striking example of another use to which it can be put.

The principles underlying these and many other electrical devices onstitute the science, or body of knowledge, of Electricity and Iagnetism. While the practical applications of this science on a irge scale are comparatively modern, its beginnings go far back ito history. In presenting a brief account of the whole subject, we hall not find it convenient to keep strictly to the historical order of liscovery of facts and principles. Some recent discoveries throw o much light on earlier observations and studies that it is desirable o take advantage of them from the outset.

ELECTROSTATICS

364. Earliest Observations.—Electrostatics is the branch of our subject which deals with the phenomena of charges of electricity that are (apart from internal motions) at rest relatively to the observer. Electricity is known to us only by its effects, and it is only with them that we are concerned, not with any vague question as to what the thing is in itself.

One of the earliest recorded observations of these effects was made about 600 B.C. by Thales, chief of the seven "wise men" of Greece. He noted that a piece of amber or jet, when rubbed with fur, acquires the property of attracting light objects, such as bits of feathers and straws. This information was handed on by other writers, but for over two thousand years practically nothing new was added to it. About 1600 A.D. William Gilbert, court physician

to Queen Elizabeth, found that many substances have the property in question; and he coined the word *electric*, derived from the Greek (elektron) for amber, to describe such phenomena. About fifty years later the word *electricity* came into use as a name for the cause of these effects. Since then discovery in this field has gone forward at an ever accelerated pace, until what was once the merest toy has become a tool of immense importance.

We now know that all solid bodies can, under proper circumstances, be electrified by rubbing with some suitable materials. The *contact* of two dissimilar substances is the essential thing; rubbing merely serves to produce closer contact. Amber is electrified by rubbing with fur, and glass is electrified by rubbing with silk. The question now arises whether the electricity on the amber and that on the glass are different in any essential respects. The following simple experiment will supply the answer.

365. Positive and Negative Electric Charges.—If one end of a rod of amber or glass is held in the hand and the part about the other end is rubbed by a suitable substance, only the part rubbed becomes electrified. Now suppose that an end, a (Fig. 228) of an amber rod A has been electrified by rubbing with fur and that the rod is then suspended by a light cord so that it is free to turn. If a second amber rod B, of which the end b has been electrified in the same way, is brought near a, as indicated in the figure, a will move away from b, showing that a and b *repel* each other. But if B is replaced by a glass rod, electrified by rubbing with silk, the electrified ends of A and the glass rod *attract* each other.

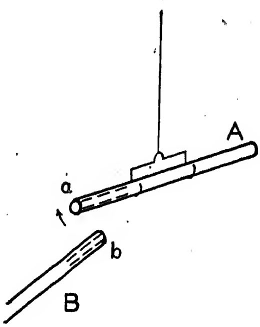


FIG. 228.

From this experiment we see that there are two different states of electrification and these we attribute to *two different kinds of electricity*. Since in many respects they exhibit opposite properties, it is convenient to call them *positive* and *negative*. Which to call positive and which negative is of little importance; but it has become the custom to call the electricity on amber electrified by rubbing with fur negative, and that on glass rubbed with silk positive. The terms positive and negative, as applied to charges of electricity, were first suggested by Benjamin Franklin, though he used them in the opposite way.

We can now use the suspended amber rod to test various electrified bodies, such as sulphur, sealing-wax, ebonite, etc. and so classify them as charged positively or charged negatively. But in each case the kind of charge depends to some extent on the rubbing substance as well as on the rubbed.

While, so far, we have been considering only the electrification of hard bodies by rubbing them with certain soft substances, such as silk or fur, it would naturally be expected that the rubbing substance would also show some electrical effects. This is found to be so, and tests by the suspended rod method show that in all cases, when one of the two things rubbed together becomes charged positively, the other acquires a negative charge.

The bodies that we test in this way consist of molecules and molecules consist of atoms, and it will now help us in describing these electrical phenomena if we review briefly what physicists have learned in the present century regarding the internal structure of *atoms,* even though the evidence for the statements must be postponed until later.

366. Atomic Structure.—An atom of any element is the smallest electrically neutral particle of it that can take part in a chemical change. There are ninety-two different elements, for example hydrogen, silver, thorium, etc. The work of the physicists has shown that an electrically neutral atom may be further broken down into positive and negative parts.

The negative part of an atom consists of minute, separate and similar parts called *electrons.* Electrons are universal constituents of matter, their mass and electric charge being the same from whatever source they are obtained. The mass of an electron is very small, about one two-thousandth of the mass of the lightest atom, hydrogen, and, in terms of grams, about 9×10^{-28} gram. Since normal atoms are electrically neutral, each must contain positive electricity, sufficient in amount to neutralize the total negative charge of all the electrons.

The positive charge, unlike the negative, is concentrated at the center of the atom in what is known as the *nucleus* of the atom. Since the electrons contribute so little to the mass of an atom, the nucleus must contain practically the whole of the mass associated with an atom. The linear size of a nucleus or an electron, so far as a definite size may be attributed to them, is very small, being about one hundred-thousandth, or 10^{-5}, of that of the atom of which it is a part.

Atoms, then, are tiny electrical structures. We may, for the present, picture them as built up from a heavy central nucleus, having a resultant positive charge, surrounded at some distance by electrons, sufficient in number to neutralize by their combined negative charge the positive charge on the nucleus. Atoms of different elements differ from one another, not only in mass or weight, but also in the number and arrangement of the electrons about the central nucleus. Further information on atomic structure as well as evidence supporting these statements will be presented later (§§449, 540, 729).

367. Conductors and Insulators.—Everyone is familiar with the fact that substances like copper and silver and, in general, all metals are better conductors of heat than amber or glass etc. In a general way substances that are good conductors of heat are also good *conductors* of electricity. This would suggest that the electron is directly responsible for both thermal conduction and electrical conduction. There are, however, difficulties in this view that have not yet been wholly overcome (§329).

Good conductors, such as metals, are substances in which the external electrons of the atoms are relatively free from the attractive force of the nucleus and can, under suitable circumstances, move through the body of the conductor. The ease with which these "free" or conduction electrons may move about in the conductor is different for different substances; and there are all gradations in conductivity from good conductors, like silver or copper, to poor conductors, like damp wood.

On the other hand, there are substances, like glass, amber, wax, etc., which are practically unable to conduct electricity and are called *insulators* or *dielectrics*. In such substances there are no free or conduction electrons, all the electrons being tightly bound in the atoms. It is possible, however, to remove an electron from an atom of an insulator. This is what occurs when, for example, a glass rod is rubbed with silk. Electrons are removed from the atoms of the glass rod and go over to the silk. The deficiency of electrons, negative charges, on the glass rod leaves it with a positive charge, while the excess of electrons on the silk gives it a negative charge. These charges are unable to move about on the silk or the glass, and remain as patches of electricity on the surface of the insulators. The process of rubbing the glass with the silk has not created electric

charges; it has merely aided in bringing about a separation of charges already present in the atoms.

From this account it is seen that the positive charge produced on the glass rod should be equal in amount to the negative charge produced on the silk. How this result is verified by experiment will be seen later (§381).

368. Coulomb's Law of Force for Electric Charges.—The force with which two charges attract or repel each other was determined experimentally by the French physicist and engineer Coulomb in 1785. His results have been verified by other investigators. If two like point charges, of amounts Q_1 and Q_2 in terms of any unit of charge, are placed r cm. apart in a vacuum, they repel each other with a force F proportional to the product of the two charges and inversely as the square of the distance between them. Hence for a vacuum and with a suitable unit of charge,

$$F = \frac{Q_1 Q_2}{r^2}$$

If Q_1 and Q_2 are of the same sign (both positive or both negative) F is positive and the force is a repulsion; whereas if they are of opposite signs, F is negative and the force is an attraction.

If the charges are on bodies immersed in some insulating medium, the force between them is less than that in a vacuum. This effect of the medium is taken account of by introducing a factor K which is known as the *dielectric constant* or *specific inductive capacity* of the medium. Thus the force between two charges Q_1 and Q_2 placed r cm. apart in a medium of dielectric constant K is

$$F = \frac{Q_1 Q_2}{K r^2}$$

From these equations it follows that the dielectric constant of a vacuum is unity. The value of the dielectric constant of air or any

VALUES OF DIELECTRIC CONSTANT K FOR VARIOUS MEDIA

Medium	K	Medium	K
Pure Water	81	Ebonite	2.72
Glass	9.9–5.5	Air	1.0006

gas at ordinary pressures is so near unity that it may be taken as unity in any but the most accurate work. We shall take it to be unity where nothing to the contrary is indicated.

It might seem that the numerical magnitude of K for any medium could be found by measuring the force between two charged bodies, first when they are in a vacuum, and next when they are immersed in the medium in question. But the medium would have to extend to a large distance compared with the distance between the charged bodies, and the procedure would be difficult and could not yield very accurate results. There are, however, other and quite accurate methods for the same purpose, but it will be necessary to postpone the account of them (§396).

369. Unit Quantity of Electricity.—Coulomb's law provides a method for defining unit charge or unit quantity of electricity by making each term in the equation $F = Q_1Q_2/Kr^2$ equal to unity. Thus, putting $F = 1$ dyne, $r = 1$ cm., $K = 1$, which means that the charges are in a vacuum, and $Q_1 = Q_2 = 1$, we get the *electrostatic unit of charge* or, briefly, 1 e.s.u. *The electrostatic unit of charge is that charge which, when placed* 1 *cm. in a vacuum from an equal and similar charge, repels it with a force of* 1 *dyne.*

The charge of an electron, as found by a method that will be described later (§400), is 4.803×10^{-10} e.s.u., so that it would require 2.08×10^9 electrons to give a total charge of 1 e.s.u.

Just as there are different units of mass, length or time, so there are other units for electric charges, namely the *electromagnetic unit of charge* or 1 e.m.u. and the *practical unit* called the *coulomb.* These units of charge will be appropriately defined later, and for the present we give merely their relationship to the electrostatic unit of charge.

1 e.m.u. of charge $= 3 \times 10^{10}$ e.s.u. of charge

1 coulomb $= 3 \times 10^9$ e.s.u. $= 0.1$ e.m.u. of charge

Both these units of charge are enormously larger than the electrostatic unit of charge. It is interesting here to note that 3×10^{10} cm. per second is the value for the velocity of light in a vacuum.

370. Electric Fields.—Coulomb's law of force for electric charges is similar to Newton's law of gravitation for particles of matter, $F = Gm_1m_2/r^2$, inasmuch as both are inverse square laws. Just as there are gravitational fields surrounding any matter, so there are electric fields surrounding any electric charges. The gravitational field of the earth may be measured by the force, g dynes, exerted on

unit mass, 1 gm. In a similar manner the electric field in any region may be measured by the force, E dynes, exerted on unit charge.

The strength or intensity E of an electric field at any point is numerically equal to the force exerted by the field on unit positive charge placed at that point. If a charge of Q e.s.u. is placed in a field of strength E e.s.u., the force exerted on the charge by the field is EQ dynes. (Compare the force on a mass m in a gravitational field which is mg dynes.)

The electric field strength at a point r cm. away from a point charge of $+Q$ e.s.u. in a vacuum or practically in air is $E = Q/r^2$ e.s.u. In a medium of dielectric constant K the field strength r cm. from a charge of $+Q$ e.s.u. is $E = Q/Kr^2$ e.s.u.

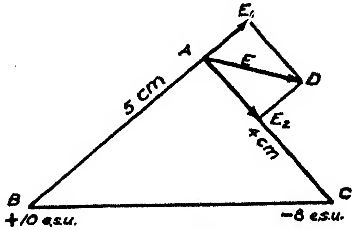


FIG. 229.

Since field strength is force per unit charge, and force is a vector quantity while charge is a scalar, it is evident that field strength is a vector, that is, it is not fully specified unless both its magnitude and its direction are given. Hence the addition of two or more field strengths must be made by vector methods.

For example, let us find the electric field strength at A (Fig. 229) due to the presence of a charge of $+10$ e.s.u. at B and a charge of -8 e.s.u. at C (in a vacuum), where $AB = 5$ cm., $AC = 4$ cm. and BAC is a right angle.

$$E_1 = +\tfrac{10}{25} = +0.40 \text{ e.s.u. directed from } B \text{ to } A$$
$$E_2 = -\tfrac{8}{16} = -0.50 \text{ e.s.u. } C \text{ to } A \text{ or } +0.50 \text{ e.s.u. } A \text{ to } C$$

The resultant of these two field strengths is found by the parallelogram law for compounding vectors and is $E = 0.64$ e.s.u. Its direction makes with AC an angle whose tangent $= 0.40/0.50 = 0.80$.

371. Electric Lines of Force.—The fact that the medium in which charges are immersed plays an important role in electrical phenomena led Faraday, an eminent English scientist of the last century (1791–1867), to introduce the concept of *lines of force* as a graphic representation of an electric field. Each of these lines of force is supposed to be drawn so that, at any point on it, the line is in the direction of the field intensity at that point, the direction of the field being indicated by an arrowhead. The idea can be illustrated by placing two charged bodies on an insulated plate and

sprinkling crystals of gypsum (hydrated calcium sulphate) on the plate. The crystals become electrified by the electric field, positive at one end and negative at the other, and gentle tapping will cause them to align themselves along the lines of force. The lines so indicated represent one section of the field of the two charges.

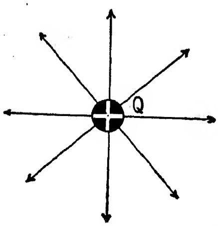


FIG. 230.

The lines of force of a point charge are radial lines diverging from the point charge (Fig. 230). Those of two unlike point charges and of two like point charges are of the forms indicated in Figs. 231 and 232. In these diagrams the lines of force end at some arbitrary points in space, and this may give rise to an incorrect idea of the properties of lines of force. In reality *one end of a line of force is on a positive charge and the other on a negative charge.* The line is a connecting link between opposite charges, following the direction of the electric field.

Faraday thought of the medium in an electric field as being in a state of stress, and he represented this stress by attributing physical properties to the lines of force. The lines in Figs. 231 and 232 present the appearance of being under tension along the lines, like elastic bands, with a mutual repulsion transverse to the lines. Attraction between two unlike charges might be thought of as due to the tension, and repulsion between two like charges as due to a sidewise pressure between the lines.

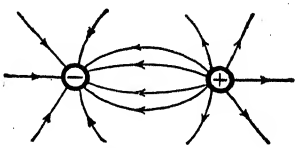


FIG. 231.

Lines of force may possibly appear to be an unnecessary addition to electrical ideas, but this is not the case. We shall see later how Maxwell, by interpreting Faraday's ideas mathematically, arrived at the electromagnetic theory of light and the possibility of creating electromagnetic waves, which ultimately led to wireless telegraphy. The electrical engineer would feel lost without the concept and language of lines of force.

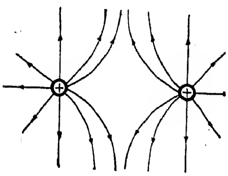


FIG. 232.

372. Lines of Force and Field Intensity.—So far we have used lines of force only in a descriptive way, to show the general nature of a field of force. A line of forc , as the term was used, can be drawn

through any point in the field, so that the number of such lines is unlimited. While each line indicates the *direction* of the field at any point on it, it tells nothing about the *intensity* of the field. We can, however, modify the device so as to represent both direction and intensity. The modification consists in using only a *limited number* of all the possible lines of force that might be drawn.

Consider first the simplest case, that of a point charge Q at O in a vacuum (Fig. 233). The field intensity at A, at a distance a from O, is Q/a^2. If a sphere S be drawn through A with O as center, all the lines of force from Q are perpendicular to it. Let us now confine attention to a limited number of the possible lines of force from O, uniformly distributed, so that at A *the number of lines of force per cm.² of surface perpendicular to them is numerically equal to the field strength there.* This means that if the field strength at A is 5 e.s.u. there are 5 of these special lines of force per cm.² (Fig. 234). Let us now restrict the term line of force to these special lines. To find the total number of them that pass through S (Fig. 233) we multiply the field strength, which is Q/a^2, by the area, $4\pi a^2$, and this gives $4\pi Q$. In this result a does not appear, and we would have come to the same conclusion if we had started the selection of lines of force to represent field strength at any other point, say B. Thus, if we draw $4\pi Q$ lines of force uniformly distributed from a charge Q, they represent, by their number per cm.², the field intensity at *any* point in the field. Confining our attention now to these lines of force, we conclude that *the number of lines of force that start from a charge Q in a vacuum is $4\pi Q$.*

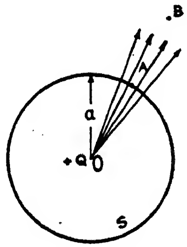


FIG. 233.

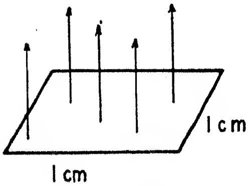


FIG. 234.

For simplicity we have considered the case of a point charge in a vacuum. Similar reasoning applies if it is in a medium of dielectric constant K; but, since the intensity at a point is then Q/Ka^2, the number of lines of force from the point charge Q is $4\pi Q/K$. In both cases the reasoning is based on the inverse square law. This method of using lines of force to represent the strength of a field would not be possible if the inverse square law were not correct.

Similar reasoning can be applied to the field of any number of point charges or any distribution of electricity. The problem is

more complex mathematically, but the result is similar. If $4\pi Q/K$ lines of force be supposed to start from each charge Q, the field strength at any point is numerically equal to the number of lines of force of the field per unit area normal to the field.

While this graphical method of representing field strength is very useful, it is not applicable if $4\pi Q/K$ is so small a number that the lines could not be thought of as uniformly distributed, and the method of direct calculation (§370) must then be used.

373. Electric Potential.—If we attempt to electrify a rod of copper or any metal by the method used with a glass rod, we shall not succeed. Charges are, it is true, produced on a copper rod by rubbing it with flannel, as may be verified by attaching an insulating handle to the rod; but when the rod is held in the hand, the charges produced pass through the hand and body to the ground.

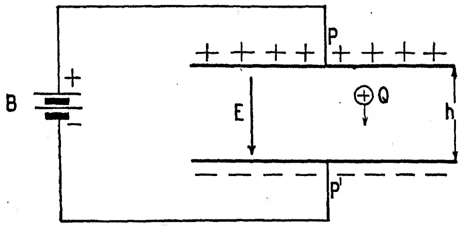


Fig. 235.

This flow of electricity is analogous to the falling of a body toward the earth. The body moves from a position of higher potential energy to one of lower potential energy (§65). If it is released with no horizontal velocity, it moves along a gravitational line of force to the earth. In a similar way, a charge of electricity in an electric field moves or tends to move along a line of force to a place where its electrical potential energy will be lower. The earth's gravitational field in a small region may be regarded as uniform, and it will be instructive to consider the analogous case of an electric field.

A uniform electric field can be produced by connecting an electric battery B to two parallel metallic plates, P and P'. For the present it is sufficient to say that the battery puts charges on the plates, positives on one, say P, and negatives on the other, P'. These charges produce a uniform electric field between the plates, the lines of force being, by symmetry, perpendicular to the plates (except near the edges). Let the field strength be E e.s.u. A charge

of Q e.s.u. placed in the field is acted on by a force of QE dynes. If it moves from P to P' along a line of force, work is done on it by the field and it loses in electrical potential energy QEh ergs, where h is the distance between the plates. To carry it back from P' to P and restore its lost potential energy would require an equal amount of work. This is analogous to the work, mgh ergs, done on a body of m grams in descending a distance h cm. in the earth's gravitational field and the equal amount of work required to raise it again.

It is convenient to have a term for expressing the tendency of electricity to flow in an electric field. The term used is *electric potential.* It may be regarded as an abbreviation for electric potential energy per unit charge, but a more exact definition is needed.

374. Potential Difference.—*The potential difference between two points in an electric field is the work done in moving a unit positive charge from one point to the other.* This definition will become clearer if we consider again the case of the two parallel conducting plates kept charged by a battery (Fig. 236). A line of force starts from a positive charge and ends on a negative charge. Suppose a positive charge to move from a to b along the line of force ab. The field does work on it and its potential energy decreases. If it is a unit charge, the work done on it by the field is Eh, where $h = ab$, and this is the decrease of potential from a to b. To carry it back from b to a would require the expenditure of work Eh by some external agent, and this is the increase of potential from b to a.

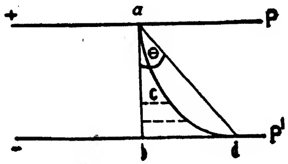


FIG. 236.

Next suppose that a unit positive charge is taken from a to d, where the straight line ad makes an angle θ with ab. The force acting on the charge is $E \cos\theta$ and the work done is $E \cos\theta \times ad$, and this is equal to Eh, since $ad \times \cos\theta = h$. It can now readily be seen that the work along a curved path acd is also Eh, since, in calculating the work, any short part of the curve is equivalent to its projection on ab.

We now see that the difference of potential between any point on P and any point on P' is the same wherever the points may be. It is therefore called the potential difference of the plates and may be denoted by $V - V'$.

As it is with differences of potential that we are concerned, we may reckon potential arbitrarily from any convenient starting-

of Q e.s.u. placed in the field is acted on by a force of QE dynes. If it moves from P to P' along a line of force, work is done on it by the field and it loses in electrical potential energy QEh ergs, where h is the distance between the plates. To carry it back from P' to P and restore its lost potential energy would require an equal amount of work. This is analogous to the work, mgh ergs, done on a body of m grams in descending a distance h cm. in the earth's gravitational field and the equal amount of work required to raise it again.

It is convenient to have a term for expressing the tendency of electricity to flow in an electric field. The term used is *electric potential.* It may be regarded as an abbreviation for electric potential energy per unit charge, but a more exact definition is needed.

374. Potential Difference.—*The potential difference between two points in an electric field is the work done in moving a unit positive charge from one point to the other.* This definition will become clearer if we consider again the case of the two parallel conducting plates kept charged by a battery (Fig. 236). A line of force starts from a positive charge and ends on a negative charge. Suppose a positive charge to move from a to b along the line of force ab. The field does work on it and its potential energy decreases. If it is a unit charge, the work done on it by the field is Eh, where $h = ab$, and this is the decrease of potential from a to b. To carry it back from b to a would require the expenditure of work Eh by some external agent, and this is the increase of potential from b to a.

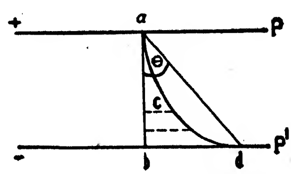


FIG. 236.

Next suppose that a unit positive charge is taken from a to d, where the straight line ad makes an angle θ with ab. The force acting on the charge is $E \cos \theta$ and the work done is $E \cos \theta \times ad$, and this is equal to Eh, since $ad \times \cos \theta = h$. It can now readily be seen that the work along a curved path acd is also Eh, since, in calculating the work, any short part of the curve is equivalent to its projection on ab.

We now see that the difference of potential between any point on P and any point on P' is the same wherever the points may be. It is therefore called the potential difference of the plates and may be denoted by $V - V'$.

As it is with differences of potential that we are concerned, we may reckon potential arbitrarily from any convenient starting-

more complex mathematically, but the result is similar. If $4\pi Q/K$ lines of force be supposed to start from each charge Q, the field strength at any point is numerically equal to the number of lines of force of the field per unit area normal to the field.

While this graphical method of representing field strength is very useful, it is not applicable if $4\pi Q/K$ is so small a number that the lines could not be thought of as uniformly distributed, and the method of direct calculation (§370) must then be used.

373. Electric Potential.—If we attempt to electrify a rod of copper or any metal by the method used with a glass rod, we shall not succeed. Charges are, it is true, produced on a copper rod by rubbing it with flannel, as may be verified by attaching an insulating handle to the rod; but when the rod is held in the hand, the charges produced pass through the hand and body to the ground.

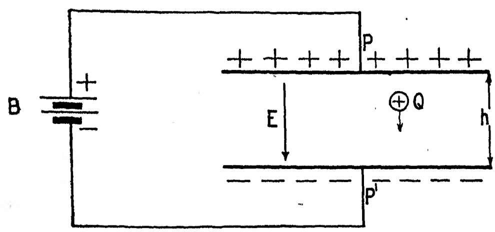


FIG. 235.

This flow of electricity is analogous to the falling of a body toward the earth. The body moves from a position of higher potential energy to one of lower potential energy (§65). If it is released with no horizontal velocity, it moves along a gravitational line of force to the earth. In a similar way, a charge of electricity in an electric field moves or tends to move along a line of force to a place where its electrical potential energy will be lower. The earth's gravitational field in a small region may be regarded as uniform, and it will be instructive to consider the analogous case of an electric field.

A uniform electric field can be produced by connecting an electric battery B to two parallel metallic plates, P and P'. For the present it is sufficient to say that the battery puts charges on the plates, positives on one, say P, and negatives on the other, P'. These charges produce a uniform electric field between the plates, the lines of force being, by symmetry, perpendicular to the plates (except near the edges). Let the field strength be E e.s.u. A charge

point taken as zero of potential, just as we arbitrarily reckon the height of a mountain from sea level. It is usually most convenient to take as zero of potential the potential of the part of the surface of the earth where the observer is, but we shall see that in some other cases a different zero has advantages (§377).

Potential or potential difference is a scalar quantity, for it is specified by a magnitude, work per unit charge, without reference to any special direction, just as level in topography does not imply a reference to any particular direction.

The term *potential gradient* is often convenient. It means difference of potential per unit distance in some direction. Thus the potential gradient along ab (Fig. 236) is $(V - V')/ab$, which is Eh/h or simply E. The potential gradient in any other direction, say along ad is $(V - V')/ad$, and, since ad is equal to $h/\cos\theta$, the potential gradient along ad is $E \cos\theta$ or the component of E in the direction of ad. Hence potential gradient in any direction means the same thing as the component of the field intensity in that direction, and it is evident that potential gradient, like field intensity, is a vector quantity.

375. Units of Potential Difference.—Any exact statement of a difference of potential requires a unit of potential difference, and this must be defined in terms of a unit of work and a unit of charge. The unit of charge used may be either the electrostatic unit or the electromagnetic unit, but the unit of work employed in both cases is the erg.

One electrostatic unit (1 *e.s.u.*) *of potential difference exists between two points when one erg of work is done by the field in moving one unit of positive charge* (1 *e.s.u.*) *from the point of higher potential to the point of lower potential.* Hence, if W is the work done in moving a charge of Q e.s.u. through a potential difference (p.d.)

$$\text{Work (in ergs)} = \text{p.d. (in e.s.u.)} \times Q \text{ (in e.s.u.)}$$

The *electromagnetic unit of p.d.* is defined in a similar way in terms of the electromagnetic unit of charge, the erg being the unit of work. Hence

$$\text{Work (in ergs)} = \text{p.d. (in e.m.u.)} \times Q \text{ (in e.m.u.)}$$

In both cases if the p.d. is $(V - V')$

$$W = (V - V')Q$$

In any concrete case, for example for the pair of plates of Fig. 235, the product $(V - V')Q$ is the same whichever unit of p.d. is used, since it is equal to W in ergs. It is therefore evident that the two units of p.d. are inversely as the corresponding units of charge. Hence

$$\text{The e.s.u. of p.d.} = 3 \times 10^{10} \text{ e.m.u. of p.d.}$$

The unit of p.d. used in practical work is called the *volt*. It is defined as being equal to 10^8 e.m.u. of p.d. Hence

$$1 \text{ e.s.u. of p.d.} = 3 \times 10^2 \text{ volts}$$

and

$$1 \text{ volt} = \tfrac{1}{300} \text{ e.s.u. of p.d.}$$

For practical purposes it is important to consider the work done when a coulomb is moved through a potential difference of a volt. Since the coulomb is $\frac{1}{10}$ e.m.u. of charge and a volt is 10^8 e.m.u. of p.d.,

$$W = 10^8 \times \tfrac{1}{10} \text{ ergs} = 10^7 \text{ ergs}$$

Now 10^7 ergs is a *joule* (§55), and the result is often stated in the suggestive form:

$$\text{Volts} \times \text{Coulombs} = \text{Joules}$$

This last form of statement is sometimes useful as an aid to the memory, and we shall use it in similar cases later.

376. Equipotential Surfaces.—An equipotential surface in an electric field is a surface over which there are no differences of potential between points on the surface. It corresponds to a level surface in a gravitational field. The surface of a charged conductor with the charge at rest is an equipotential surface; for if it were not, charges would be moving from places of higher to places of lower potential, so that the electricity would not be at rest.

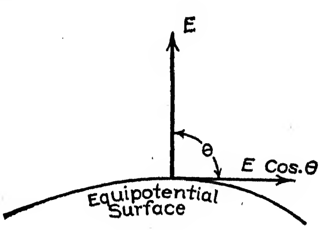


FIG. 237.

At points on the surface of a charged conductor lines of force begin or end. Where they start from or end on the surface they must be at right angles to the surface, if the charge on the conductor is at rest. If they were in a direction inclined to the surface, the

field intensity would have a component $E \cos \theta$ parallel to the surface, and the charge on the conductor would not stay at rest. Thus in Fig. 235 the surfaces of the plates are equipotential surfaces; they are horizontal and the lines of force are vertical (except near the edges of the plates).

The term equipotential surface is not confined to the surfaces of conductors. Any geometrical surface supposed to be drawn so that at each point it is perpendicular to the line of force through the point is an equipotential surface. If it were replaced by a conductor, there would be no components of intensity tending to cause a flow of electricity along the surface, that is, there would be no potential differences on the conducting plate. In Fig. 235 any plane surface between the plates and parallel to them is an equipotential surface.

377. Potential Due to a Point Charge.—So far in discussing potential we have considered only uniform fields of force. As a step toward finding the potential in a field that is not uniform, let us consider the case of a positive charge Q on a body of such small size that Q may be regarded as concentrated at a point O. The potential V decreases with increase of distance r, since a positive test charge would tend to move away from O. The problem is to find how V depends on r. It can be shown that, for a field in a vacuum (or practically in air), if we take the potential at a great distance from O (at "infinity") as zero potential,

$$V = \frac{Q}{r}$$

For let P_1 and P_2 (Fig. 238) be two points on a line of force from O, their distances from O being r_1 and r_2, and suppose $(r_2 - r_1)$ to be

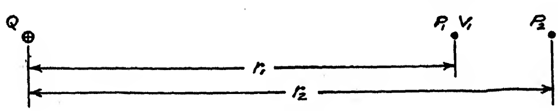


FIG. 238.

very small compared with r_1 or r_2. The field strength at P_1 is Q/r_1^2 and at P_2 it is Q/r_2^2, and we may take Q/r_1r_2 as its value in the short step P_1P_2. Then, equating the decrease of potential to the work done by the field on unit charge going from P_1 to P_2, we get

$$V_1 - V_2 = \frac{Q}{r_1 r_2}(r_2 - r_1) = \frac{Q}{r_1} - \frac{Q}{r_2}$$

This equation means that the decrease of potential is equal to the decrease in the value of Q/r, and this cannot be so unless the potential at the distance r is Q/r, with the possible addition of a constant. If, however, we take the potential "at infinity" as zero potential, V is zero when Q/r is zero, so that there can be no constant to be added. We thus get $V = Q/r$. Potential reckoned from potential "at infinity" as zero is sometimes called absolute potential.

In the preceding we have supposed Q to be a positive charge. If it is negative, V in the equation also becomes negative, which means that potential decreases from infinity up to the charge.

If the field is in a dielectric of dielectric constant K, we should have to start with Q/Kr^2 for the field strength at a distance r, and this would have given Q/Kr as the potential at a distance r.

Having now found an expression for the potential due to a single point charge, we can get the potential due to a number of point charges by simple addition, since potential is a scalar quantity (§374).

The reader who is familiar with the integral calculus will readily see that, when we take the potential at infinity as zero potential, the potential V at a distance r from a positive charge Q in a medium of dielectric constant K is given by

$$V = \int_r^\infty \frac{Q}{Kr^2}dr = \frac{Q}{Kr}$$

378. Charged Spherical Conductor.—If a spherical conductor carrying a charge Q is sufficiently far from other bodies, the charge on it distributes itself uniformly over the surface. The $4\pi Q$ lines of force from the charge (§372) diverge symmetrically from the sphere, as if they originated at its center (Fig. 239). Hence the field strength and potential at the surface and outside are *the same as would be caused by a charge* Q at the center.

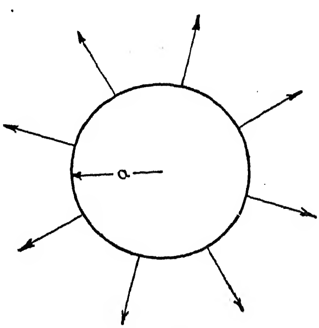


FIG. 239.

From this relation it follows at once that, if the radius of the sphere is a and it is in a vacuum, the field strength at its surface is Q/a^2 and the potential is Q/a; and, at any point outside the sphere, at a distance r from the center, the field strength E is Q/r^2 and the

3. The earth connection is now broken, C being still present. This leaves AB with an excess of electrons at A (Fig. 243).

4. Finally C is removed. The electrons then distribute themselves over the entire surface of AB, which remains charged negatively (Fig. 244).

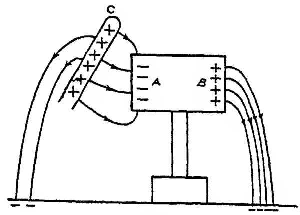


FIG. 245.

Let us now consider what happens from the point of view of lines of force and potential. Lines of force diverge from the charge Q on the rod, the number of them being $4\pi Q$. Some of them fall on A (Fig. 245) and induce a negative charge at A, and an equal number of lines of force then extend from B to the ground. When B is earthed the lines of force from B to the ground disappear (Fig. 246).

As regards potential, AB is initially at ground potential. The approach of the positively charged rod raises AB to a positive potential, the same at all points. Earthing B reduces AB back to zero potential, the positive potential due to C being neutralized by the negative potential due to the negative charge acquired by AB. When the earth connection is broken, AB is still at zero potential; but when C is removed AB remains charged negatively and at a negative potential.

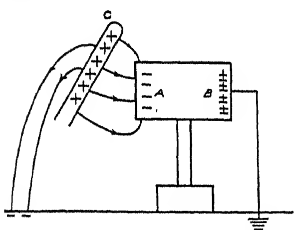


FIG. 246.

The question sometimes arises whether it makes any difference what point on AB is grounded. It does; for example, if A is grounded (Fig. 247) some lines of force that went to A are diverted to the wire, and the resultant negative charge on AB will be smaller than when B was earthed.

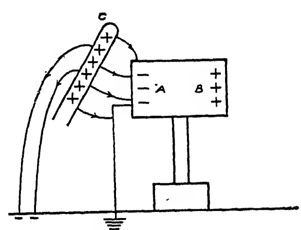


FIG. 247.

In its final charged condition AB possesses a store of energy due to its charge and potential; it could raise a light positively charged body from the ground. The source of this energy is the work that has to be done to separate the positively charged rod from the negatively charged conductor.

potential V is Q/r. The way in which E and V decrease with distance is shown graphically by the curves in Fig. 240.

Inside a charged spherical conductor, whether it is solid or hollow, the field strength is zero and the potential is that of the surface.

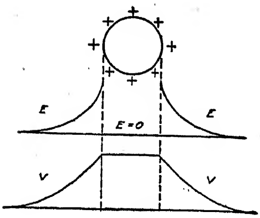


FIG. 240.

For if the sphere is solid, there are no potential differences in a conductor (in electrostatics); hence there are no lines of force and the field strength is therefore zero. If the sphere is hollow and does not enclose a charge and there is no charge on its inner surface (§384), there are no lines of force inside it, for they would have to begin and end on opposite charges.

379. Electrostatic Induction.—It is possible to charge a conductor by bringing a charged body near to it, but not into contact with it. This is called *electrostatic induction*. The process may be described as follows:

1. A positively charged rod C is brought near the end A of a conductor AB that is on an insulating stand (Fig. 241). Negative

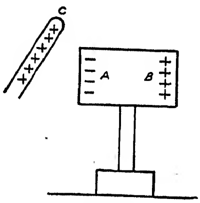


FIG. 241.

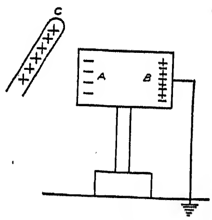


FIG. 242.

electricity is *induced* on the end A and positive on B. Conduction electrons are attracted by C; if C is removed they flow back.

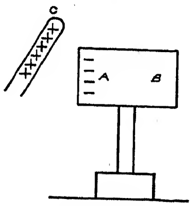


FIG. 243.

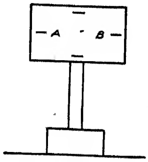


FIG. 244.

2. While C is present, AB is connected by a wire to the ground (Fig. 242); electrons flow from the earth and neutralize the charge on B.

380. Electroscopes and Electrometers.—Up to the present the only method we have used for finding the sign (+ or −) of a charge has been attraction or repulsion of a rod with a charge of known sign. A much more sensitive device is the *gold-leaf electroscope* (Fig. 248). Two gold leaves L, attached to a conducting rod R and cap C, hang in a conducting vessel D, but are insulated from D by a plug E of amber or sulphur. D may be of metal with a window for viewing the leaves; or it may be of glass, coated wholly or partly on the outside with metal foil, so that it is in effect a conductor. Leaves of gold are used because of the extreme thinness to which gold leaf can be rolled.

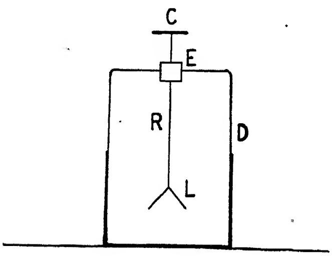


Fig. 248.

The outside of D is kept at some definite potential, usually that of the ground. The leaves can be charged either by conduction (touching C with a charged rod) or by induction. Let us suppose that they have a positive charge and are therefore at a positive potential. On D there is a negative charge, due to induction by the charge on the gold leaves. Between the leaves and D there is an electric field. The leaves spread apart because of attraction between the charge on them and the charge on D. The following simple experiment will be instructive in this connection (Fig. 249). The electroscope is placed on an insulating stand and the cap C is connected to D by a wire. A charge is then given to the electroscope, but the leaves do not diverge. If the wire is removed by an insulating rod, there is still no divergence. But if D is now grounded, the leaves diverge, for there is now a field between the leaves and D.

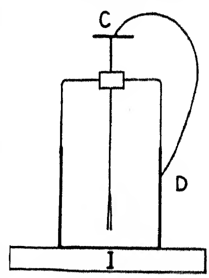


Fig. 249.

An electroscope can be used to find the sign of a charge on a body. When the electroscope has been charged positively, D being grounded, a positively charged rod, brought near C, makes the leaves diverge more widely. If the rod be charged negatively, the divergence of the leaves will decrease. The electroscope can also be used to indicate roughly the magnitude of a potential difference. If C be connected to a body at one potential and D to a body at a

different potential, the leaves will diverge more widely the greater the potential difference.

An *electrometer* is an instrument for measuring a potential difference. The Braun electrometer (Fig. 250) may be regarded as a modification of the gold-leaf electroscope, one gold leaf being replaced by a vertical metallic plate and the other by a metallic needle or indicator that can rotate about a horizontal axis a little higher than the center of gravity of the needle. It is provided with a vertical graduated circle on which potential differences can be read. It is useful for measuring potential differences of the order of 1000 volts.

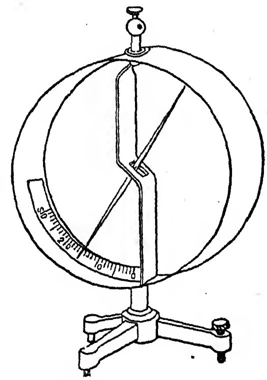


Fig. 250.—Braun electrometer.

The *quadrant electrometer* (Fig. 251), due to Lord Kelvin, is useful for measuring very small potential differences. The "needle" is a very light paddle-shaped metallic disk suspended by a very fine wire. It hangs in a shallow cylindrical box that is divided into four quadrants A, B, C, D mounted on amber supports. Diagonally opposite quadrants are connected in pairs by wires. The needle is connected to a battery and kept at a constant potential of the order of 100 volts. A potential difference is measured by making connections to the two pairs of quadrants and reading the deflection of the "needle"

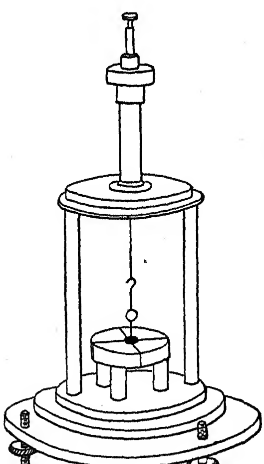


Fig. 251.—Quadrant electrometer.

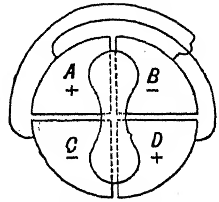


Fig. 252.—Quadrants.

by the mirror and light method. The instrument can be used for potential differences as small as 0.005 volt. A. H. and K. T. Compton increased its sensitiveness further by having one quadrant

adjustable vertically and giving the needle a small tilt. For high or alternating p.d.'s the needle is connected to one pair of quadrants.

381. Experiments with an Electroscope.—When a positive charge is produced in any way, an equal negative charge is also produced. Faraday showed this to be true in the case of charging by induction by using an electroscope and a metal vessel connected to it. He found a metal ice pail convenient for the purpose, and the experiment is usually referred to as Faraday's "ice-pail experiment"; but any light metal vessel *B* will serve for the purpose. If it is placed, uncharged, on the cap *C* of an electroscope (Fig. 253) and a positively charged body *A* is lowered into it by a silk thread, the leaves will diverge and become steady when *A* is well within *B*. If *A* be now moved around within *B*, but not touching it, the divergence of the leaves will not change; and no change of the divergence will take place if *A* is allowed to touch the inside of *B* and is then lifted entirely out of *B*; but *A* will then be found to be wholly uncharged.

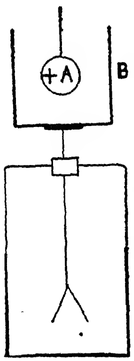


FIG. 253.

When *A* with its positive charge is well within *B*, all the lines of force from *A* end on *B* (Fig. 254) and give it a negative charge, an

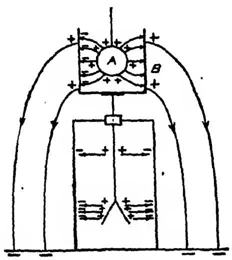


FIG. 254.

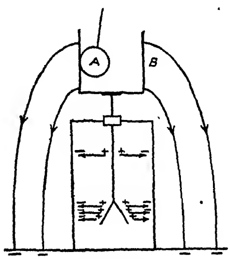


FIG. 255.

equal positive charge going to the outside of *B* and the leaves of the electroscope. As there was no change in the divergence of the leaves when *A* touched *B*, the inducing positive charge on *A* and the induced negative charge on *B* must have been equal, so that they just neutralized each other, leaving the lines of force outside *B* unchanged (Fig. 255).

The same arrangement can be used to show that equal and opposite charges are produced in electrification by rubbing (Fig. 256). If a piece of flannel is wrapped on one end of an amber rod and held

by a silk thread, so that the rod can be rotated inside the flannel, any charges that may be produced do not escape. If the rod and flannel, while together, are thrust into B, no divergence of the leaves is observed; but if they are separated and tested separately, equal divergences are obtained. Thus equal quantities of electricity of opposite signs were produced.

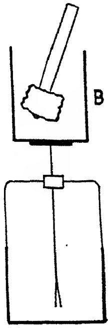


FIG. 256.

Faraday's arrangement can also be used to prove the statement made earlier (§364) that, when electrification is produced by rubbing or friction, the only function of the rubbing or friction is to produce more intimate contact. For this purpose B is partly filled with pure water (Fig. 257), and a glass vessel D coated with paraffin wax is placed in the water. When D is pulled up and out of the water, the electroscope leaves diverge with a positive charge, D being charged negatively. There is no true friction between the water and the paraffin, but there is intimate contact, and electrons pass from the water to the paraffin. There is reason to believe that between any two dissimilar substances in contact there is a contact potential difference. In a "contact layer," roughly atomic in depth (about 10^{-8} cm.), electrons are transferred from one substance to the other until, if either is a non-conductor, the potential difference is neutralized. What happens if both are conductors will be considered later (§466).

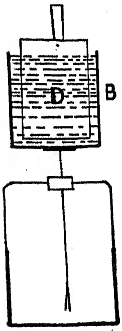


FIG. 257.

It will now be seen that "production" of electricity does not mean creation of electricity. It rather means extraction, in the sense in which gold is extracted from an ore. In some cases the electrons extracted were free or conduction electrons, in other cases they were dragged or knocked out of atoms.

382. Distribution of Charge on a Conductor.—Although the surface of a charged conductor (the electricity being at rest) is an equipotential surface, the surface density, or charge per unit area, may be far from uniform. It can be studied experimentally by means of a "proof plane" and an electroscope. A proof plane is a small metal disk on an insulating handle. When the disk is placed anywhere on a charged conductor, it receives a charge proportional to the surface density there, the sign and magnitude of which can be found by the electroscope. The results for two forms of con-

ductor are indicated in Figs. 258 and 259 by the distribution of the + signs. Evidently the density of electrification is *greatest at places of greatest curvature*, though the potential is the same everywhere over the surface.

A partial explanation of this result is found in the relation between the curvature $1/r$ and the surface density σ in the case of an

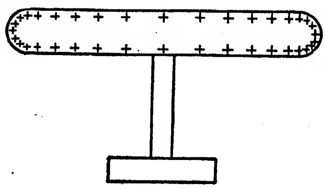


FIG. 258.

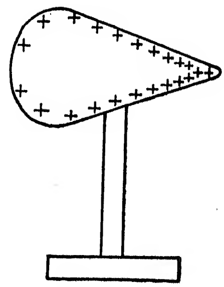


FIG. 259.

isolated spherical conductor that is charged to a given potential V(abs.). If the whole charge on it is Q,

$$V = \frac{Q}{r} \quad \text{and} \quad \sigma = \frac{Q}{4\pi r^2} \quad \therefore \sigma = \frac{V}{4\pi r}$$

Thus, for a given potential, the surface density is proportional to the curvature. This result would apply, at least approximately, to a "body" consisting of a large sphere and a small sphere joined by a long wire, provided the spheres are so far apart that induction by one on the other may be neglected. But if the spheres are closer together or in direct contact, while the surface density is still greatest where the curvature is greatest, one is not simply proportional to the other.

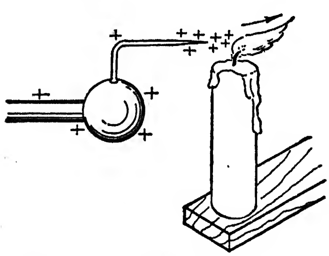


FIG. 260.—Candle flame blown out by discharge of electricity from a pointed wire. From Foley's Physics.

383. Discharge from Points.—If the potential and surface density of a pointed part of a conductor are sufficiently high, leakage from the point takes place; the adjacent air becomes charged and, being repelled, forms a miniature wind that can blow a candle out (Fig. 260). There is also a mechanical action on the pointed conductor, as is illustrated, together with

the law of action and reaction, by the "electric whirligig" (Fig. 261). It rotates when kept charged by a generator.

This discharge from points and some of its effects have been known for a long time, but the closer study of its details form an interesting chapter in modern Physics. The brief sketch of it that follows will be amplified later.

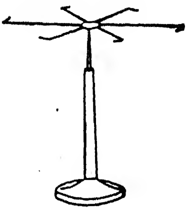


FIG. 261.

Air, under ordinary conditions, contains about 3×10^{19} molecules per cc. In this enormous number there are always a few thousand charged molecules, called *ions*. An ion is a molecule that has lost or gained one or more electrons, chiefly through the action of radioactive materials in the walls of a room or in the ground and by cosmic rays. The process of forming ions is called *ionization*. About five pairs of ions (a positive and a negative) are produced in this way, every second, in each c.c. of air, three by radioactivity and two by cosmic rays. The total number of ions in the air at any time is an equilibrium number, as many being removed by recombination as are produced by ionization.

Near a pointed part of a charged conductor there is an intense field in the air (Fig. 262). The charge on the point attracts ions of opposite charge and repels ions of like charge. Both kinds may attain sufficient velocity and energy to ionize neutral molecules by impact, thus producing more ions, which may again give rise to still others, and so on, the process being cumulative. If it goes far enough, it gives rise to a faint visible light and the discharge is then called a *corona* discharge.

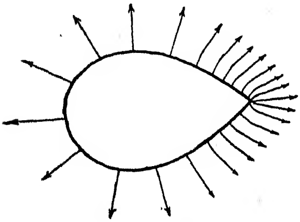


FIG. 262.

With a sufficiently intense field very high ionization is produced; and then a sudden discharge, visible as a spark or flash, may take place to some other conductor. This is the familiar phenomenon of lightning, the discharge being from cloud to cloud or from cloud to earth. In this case there are charged clouds due perhaps to the breaking up of water droplets in a swift vertical air current. These charged clouds induce charges of opposite sign on the earth or neighboring clouds, producing an electric field between them. If there is a well grounded pointed conductor below the cloud, such as a lightning rod, then the point

with its intense field may be able to discharge ions quietly in air. This reduces the electric field and may prevent a direc ning stroke. But if its action is not sufficiently rapid or if it well earthed, it may "attract" a discharge. When the dis takes place, heat is produced along the path of the discharg air expands causing, for a moment, a partial vacuum; the sur ing air rushes in and thus gives rise to thunder. In the absen lightning rod, a tree or even a head may play a similar part. in a thunderstorm it is not wise to stand under or to stand up in an open field.

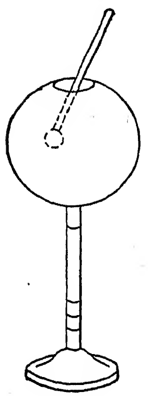


FIG. 263.

384. Conditions within a Hollow Conductor.- inner surface of a charged hollow conductor c tested for charge by passing a proof plane thro small opening in the conductor (Fig. 263). C tests show that, if there is no insulated charge cavity, *there is no charge on the inner surfac* matter how great the charge on the outer surfac be. From this an important conclusion can be d. Since there is no charge on the inner surface can be no lines of force in the cavity, for lines of must have their ends on charges. As there a lines of force, there is no field intensity. Henc potential is the same everywhere in the cavity and therefore to the potential of the conductor itself.

This property of a closed conductor is the basis of *electro shielding*. Faraday covered a large box with a netting of cc wire and went inside it with an electroscope. When the net was charged, even to the point of sparking from corners, he fe electrical effects, and the leaves of the electroscope did not div Wires leading to a delicate electrometer, used for measuring small charges or potentials, are usually enclosed in earthed l tubes, with amber as an insulator between wire and tube, for pr tion against electrical disturbances in the vicinity.

385. Evidence for Inverse Square Law.—The fact, establi by experiment, that there is no field in a cavity in a charged ductor affords a more rigorous proof of Coulomb's inverse square than a direct measurement of the force between two charges (§ provides.

For simplicity consider a spherical shell of conducting mate and suppose it to be charged with positive electricity, the ch

being uniformly distributed over the surface. Let P be any point within the shell. Imagine the whole surface of the shell to be divided up by double cones of small angles with a common vertex at P. Let the areas of the bases of one such double cone be S_1 and S_2, and let the charges on them be Q_1 and Q_2. Then, since the surface density is uniform,

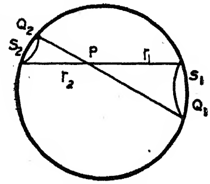


FIG. 264.

$$\frac{Q_1}{Q_2} = \frac{S_1}{S_2}$$

It is known, from geometrical principles, that the bases are equally inclined to the common axis, and, if their distances from P are r_1 and r_2,

$$\frac{S_1}{S_2} = \frac{r_1^2}{r_2^2}$$

Hence

$$\frac{Q_1}{Q_2} = \frac{r_1^2}{r_2^2}$$

and therefore

$$\frac{Q_1}{r_1^2} = \frac{Q_2}{r_2^2}$$

So far no particular law of force has been assumed. If, however, the inverse square law is correct, Q_1/r_1^2 is the magnitude of the field intensity at P due to Q_1 and Q_2/r_2^2 that due to Q_2. Moreover these field intensities are in opposite directions. Since they are equal and opposite, they neutralize each other; and the same is true for each of the double cones that fill the whole space. Hence the field intensity within the shell is zero, *if the inverse square law is true.* But if the true law were something different, say an inverse cube law, the field intensity in the cavity should evidently not be zero.

If it could be shown by experiment that the field strength in the cavity is exactly zero, it would be a proof that the exponent in the $1/r^2$ law is exactly 2. To test the question Cavendish (1773) used two concentric shells and examined the inner for a charge after the outer one had been charged, connected to the inner one, then disconnected and discharged; he found none detectible by his pith-ball electrometer. Maxwell, a century later, repeated the test with a quadrant electrometer and concluded that the exponent

was 2 to within at least 1 part in 20,000. Recently Plimpton and Lawton, using modern apparatus, have shown that the exponent is 2 to within 1 part in 1,000,000,000.

386. Volta's Electrophorus.—This is a convenient device for getting a small charge by induction. A plate of resin, A, (Fig. 265) that rests in a shallow metal vessel S is first charged negatively by rubbing or beating with flannel. A metal disk B, held by an insulating handle H, is placed on A. A positive charge is induced on the lower surface of B and a negative charge on the upper surface. If B is now earthed by touching it with a finger, it is left with a positive charge and at zero (earth) potential. When it is raised by the insulating handle H, its potential becomes positive and it acquires energy equal to the work required to separate it from A.

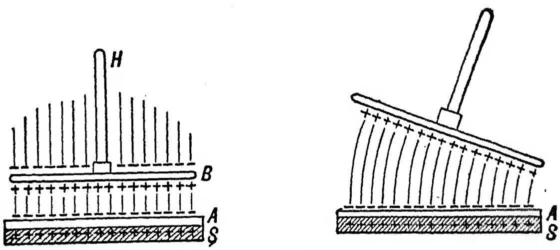


FIG. 265.

While the charge on B is not large, the potential is quite high, and a spark will pass to a finger brought near it.

The process of charging and discharging B can be repeated many times without much loss of charge by A. The contact between A and B is only at a few points, so that very little negative electricity goes to B.

Later a series of small disks on a revolving glass plate was used, and this again developed into a number of electrostatic generators. The action of any one of these is best studied by examining it. We shall describe here only one of the most common of the older type and then the latest and most efficient one.

387. Electrostatic Generators.—In the *Toepler-Holtz generator* in its simplest form (Fig. 266) there is a stationary glass plate A that carries thin metallic strips s and s' and, in front of it, a rotating glass plate B, on which there are small metallic discs or buttons $b, b, \ldots$. As two opposite buttons enter the regions in front of s and s', they are touched by brushes of flexible wire attached to s and s', and,

as they leave these regions, they are touched by similar brushes on a stationary metallic rod r. The action of the generator starts with some slight charge on s or s', due to friction of a brush on a button, or, if necessary, to some outside source of charge. Suppose that s has originally a slight negative charge. Two buttons in contact with r then become oppositely charged by induction, and, as B revolves, they touch the brushes attached to s and s' and share their charges with s and s'. Thus charges on s and s' are built up as larger charges are induced on bb, until they begin to draw discharges from metallic combs c and c' with sharp points. This

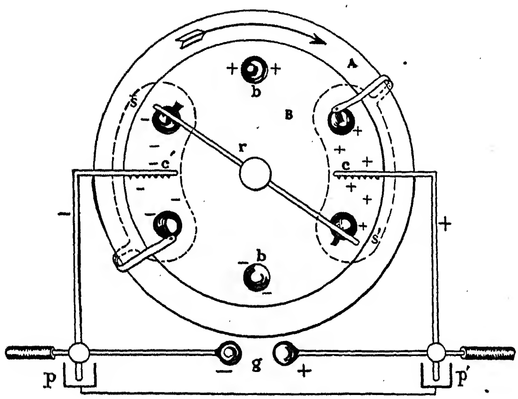


FIG. 266.

charges two condensers P and P', the action of which in storing a charge will be considered presently (§388). When the potentials of the condensers become sufficiently high, discharge takes place across the spark-gap g. This does not affect the charges on s and s', so that recharging of the condensers immediately takes place.

In the *Van de Graaff generator* (Fig. 267) a silk insulating belt S passes over two pulleys and is kept in motion by a motor that drives the lower pulley. The other pulley is inside a large hollow metal sphere, the belt passing through a hole in the sphere. A, C, D, F, G are metallic combs of sharp points. B and E are metallic plates connected by wires to A and F respectively. D and G are connected to the sphere by wires. C is grounded. Suppose AB has initially a positive charge, supplied by an auxiliary generator or due to rubbing of the belt on the pulleys. The charge on B acts inductively on C, drawing from C a stream of electrons that are "sprayed"

on the belt. As the belt with its negative charge enters the sphere, the field between the belt and the points F causes a discharge of the electrons to F, thus charging E negatively. The points G also "collect" electrons from the belt and transmit them to the sphere, so giving it a negative potential. The belt, now practically uncharged, passes over the pulley and between E and D, where the

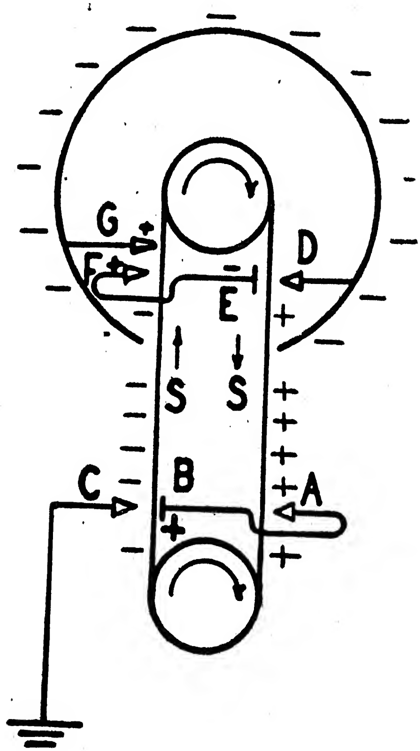


FIG. 267.

charge on E causes D to discharge positive electricity to the belt, the negative going to the sphere and increasing its negative potential. At A positive electricity is collected from the belt and reinforces the charge on AB. The cycle is thus complete, and the net result is a steady transfer of electrons from the earth to the sphere. Machines of this kind have produced a potential of 10^7 volts when discharging 10^{-5} coulombs per second. The mechanical power of

this discharge is $10^7 \cdot 10^{-5} = 100$ joules per second, that is, 100 watts or about $\frac{1}{7}$ of a horsepower.

388. Electric Condensers. Capacity.—A condenser is a device for storing a large charge without a great rise of potential. An experiment will illustrate its principle.

A is a conducting plate that is earthed (Fig. 268a) and B is a similar plate placed parallel to A and connected to an electroscope. As regards electrical relations, we may consider B, the electroscope and the wire connecting them as a single conductor; we may call it the B-system. If a charge be given to B, the divergence of the leaves of the electroscope will be a measure of the potential of the B-system relatively to the earth as zero, for it will depend on the field between the leaves and the earthed casing of the electroscope.

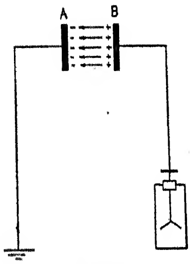


FIG. 268a.

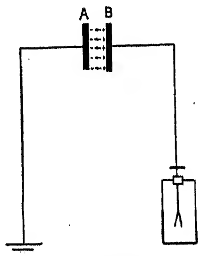


FIG. 268b.

Now suppose A to be moved toward B (Fig. 268b). The divergence of the leaves of the electroscope will decrease, indicating a fall in the potential of the B-system. This may be considered as due to the approach of the negative charges on the plate A when it is moved towards B. To understand this in a quantitative manner, consider the field between A and B. If the field strength is E, the distance between the plates d and the potential difference of the plates V,

$$V = Ed$$

Moving A toward B does not change the charge Q on the B-system; and if we suppose B to be of large size compared with the leaves of the electroscope, the charge on B will be practically unchanged. If the plates are already so close together that all of the $4\pi Q/K$ lines of force from B (§372) fall on A, the number of lines of force per cm^2 in the field will not change as A is moved toward B, so that the

field intensity E may be regarded as constant. Hence V is proportional (at least approximately) to d. Thus the presence of A, kept at a constant potential by its connection to the earth, keeps B at a moderate potential, even when a very large charge is given to it.

The two plates A and B constitute what is called a *condenser*. Its capacity is defined as: *the magnitude of the charge on either plate divided by the magnitude of the difference of potential of the plates.* Hence if the charge is Q and the potential difference V, and if we denote capacity by C,

$$C = \frac{Q}{V} \qquad \text{and} \qquad Q = CV$$

Here and in similar cases later we denote the difference of potential of the plates of a condenser by V instead of $(V - V')$. This is often convenient for brevity, where the meaning is clear.

389. Units of Capacity.—In stating the equation $C = Q/V$ by which capacity was defined, no particular units were mentioned, though it was, of course, implied that C should be 1 if Q were 1 and V also 1. This, in reality, supplies a definition of unit capacity in terms of any pair of units of charge and potential. Whatever the units of Q and V may be, *unit capacity is the capacity of a condenser such that unit charge produces unit potential difference of the plates.* It will be sufficient here to consider two such units of capacity.

The *electrostatic unit of capacity* is defined in terms of the e.s.u. of charge and the e.s.u. of potential. The *farad* (so named in honor of Faraday) is defined in terms of the coulomb and the volt. The relation between these two units is found by applying the equation $C = Q/V$. Thus, with the aid of §§369, 375, we get

$$1 \text{ farad} = \frac{1 \text{ coulomb}}{1 \text{ volt}} = \frac{3 \times 10^9 \text{ e.s.u. of charge}}{\frac{1}{300} \text{ e.s.u. of p.d.}}$$
$$= 9 \times 10^{11} \text{ e.s.u. of capacity.}$$

The farad is much too large for practical purposes. A microfarad is the usual unit. It is one millionth, 10^{-6}, of the farad and is sometimes denoted by μf. Even this unit is at times too large for convenience, and then one thousandth of it, denoted by mμf, or even one millionth of it, denoted by $\mu\mu$f, is used.

390. Dielectric in a Condenser.—In the arrangement of Fig. 268 suppose the plates A and B to be in their original positions, and

let a glass plate be placed between A and B (Fig. 269). Again the divergence of the leaves of the electroscope decreases, indicating a fall of the potential of B with no change of charge, and therefore an increase of the capacity of the condenser AB. Similar results are obtained when plates of other non-conductors are used.

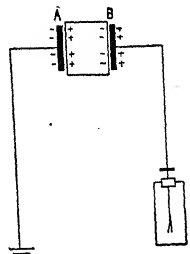


Fig. 269.

These effects of the dielectric in a condenser can be explained by the *polarization theory* of the dielectrics. Each atom in a dielectric consists of a positive and a negative part. In an electric field these parts are displaced relatively to each other to an extent proportional to the field strength, so that the atom becomes a dipole (Fig. 270) and is described as *polarized*. The result is a layer of positive charges on the surface of the dielectric that faces A and a layer of negative charges on the other face. These surface layers produce a field opposed to the field of the charges on A and B, so that the resultant field from B to A is decreased and the potential difference between B and A is reduced. Since the charge Q is unchanged, while the potential difference, V, of the plate is reduced, the capacity, Q/V, is increased by the presence of the dielectric.

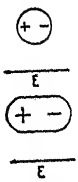


Fig. 270.

This explanation of the effect of a dielectric evidently applies also to the attraction between two oppositely charged bodies, and it will now be seen that the constant K in Coulomb's law must come into any expression for calculating the capacity of a condenser from its dimensions and the dielectric between the plates.

391. Parallel Plate Condenser.—A parallel plate condenser consists of two similar parallel plates P_1 and P_2 (Fig. 271) with some dielectric between them. Denote the distance between P_1 and P_2 by d in cm, the area of each by A in cm.^{2}, and the dielectric constant by K. If there is a charge $+Q$ e.s.u. on P_1 and a charge $-Q$ on P_2, the whole number of lines of force from P_1 to P_2 is $4\pi Q/K$ (§372) and the number per cm.^{2} is $4\pi Q/KA$. This is therefore the intensity, E, of the field in the dielectric, in e.s.u. If the potential difference of the plates is V e.s.u., then $V = Ed$. Hence

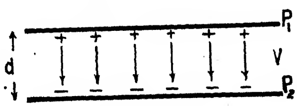


Fig. 271.

$$V = Ed = \frac{4\pi Qd}{KA}$$

Thus the capacity of the condenser is

$$C = \frac{Q}{V} = \frac{KA}{4\pi d}\text{ e.s.u.}$$

Very compact condensers of large capacity are made by using a large number of sheets of tinfoil separated by mica or some other dielectric (Fig. 272). Alternate sheets are connected together to form one group, which may be regarded as one plate of the condenser, and the others are connected to form a second group, constituting the second plate. A laboratory form of such a condenser is shown in Fig. 273.

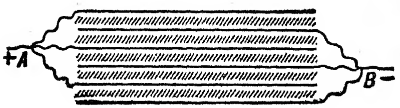


FIG. 272.

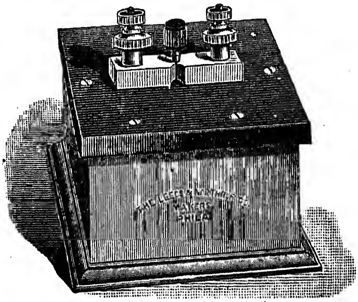


FIG. 273.

Variable air-condensers (Fig. 274) are used for "tuning" radio instruments. In Fig. 274*a* the plates A and B are in the position of maximum capacity, while in Fig. 274*b* the capacity is a minimum.

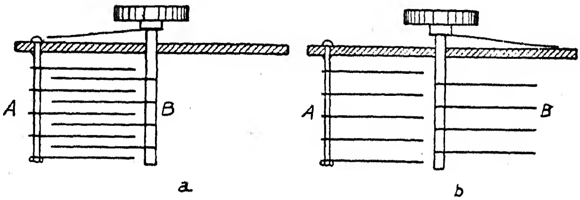


FIG. 274.

The Leyden jar (Fig. 275) is of historical interest, as it was the first practical form of condenser. It consists of a glass jar coated inside and outside to about two thirds of its height with tinfoil. When it has been charged to a high potential difference of its plates, it gives a sharp bright spark on discharge. As the energy of the discharge is sufficient to give a violent shock, discharging tongs on an insulating handle are used (Fig. 276). In the discharge there is one feature of great interest that we shall consider later (§518).

392. Spherical and Cylindrical Condensers.—A conducting sphere A, within a concentric conducting shell B (Fig. 277) and some

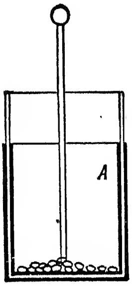


FIG. 275.

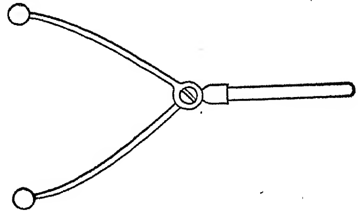


FIG. 276.

medium between them as dielectric constitute a spherical condenser. Let the radius of the sphere be a cm. and the inner radius of the shell b cm. The condenser is usually charged by earthing B and charging A by means of a wire that is inserted through an aperture in B. If there is a charge $+Q$ e.s.u. on A, all the lines of force from A end on B and there is a charge $-Q$ on B.

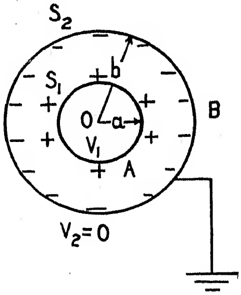


FIG. 277.

The potential difference $V_1 - V_2$ between A and B is the work in ergs done by the field on unit positive charge taken from A to B. Now the field strength within B due to the charge on B is zero (§378). Hence the field between A and B is due to the charge $+Q$ on A. Since the lines of force of $+Q$ are radial, as if the charge were at the center O, to find the potential it produces outside A we proceed exactly as in §377. We thus get

$$V_1 - V_2 = \frac{Q}{Ka} - \frac{Q}{Kb} = \frac{Q}{K}\left(\frac{b-a}{ab}\right)$$

Hence

$$C = \frac{Q}{V_1 - V_2} = K\frac{ab}{a-a} \text{ e.s.u.}$$

We have supposed that, in charging the condenser, the shell B was earthed, and this is the usual way. If the sphere were earthed and the shell charged, there would be a charge on its outer surface

as well as on its inner surface, and the capacity would not be given exactly by the above formula.

A cylindrical condenser consists of a cylinder surrounded by a cylindrical shell with a dielectric between them (Fig. 278). If the

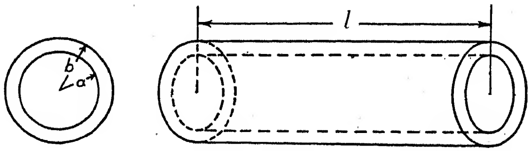


FIG. 278.

radii are a and b and the length l is large compared with a or b

$$C = K\frac{l}{2\log_e \frac{b}{a}}$$

but the mathematical proof must be omitted.

393. Capacity of a Conductor.—The expression for the capacity of a spherical condenser, with the shell or outer plate earthed, can be written in the form $\dfrac{Ka}{1 - \dfrac{a}{b}}$. Let us now suppose that b is so large compared with a that we may neglect a/b compared with 1. The expression then becomes

$$C = Ka$$

This result may be obtained very simply from the results of §378. Ka still stands for a capacity, but this capacity is a property only of the sphere and the medium in which it is. Moreover the medium is supposed to extend to a great distance compared with the radius of the sphere, without any other objects intervening. Such a sphere is usually described as "isolated" electrically, and Ka is called the capacity of the sphere when in the medium of dielectric constant K. Unless the sphere is very small, the only medium of sufficient extent usually available is the atmosphere; and the value of K for air (1.0006) is very nearly 1. Hence the capacity of the sphere in e.s.u. may, for practical purposes, be taken as equal to its radius, a, in cms.

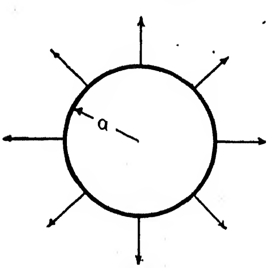


FIG. 279.

Evidently we can extend the same reasoning to any insulated conductor that is so far from other bodies that it may, for the purpose in view, be regarded as "isolated." It has a capacity that depends only on its form and dimensions and the medium, though there may be no simple formula for calculating it from them. We can, however, define it by

$$C = \frac{Q}{V}$$

where V is the potential to which the conductor is raised from zero by a charge Q given to it.

An insulated conductor that is not "isolated" also has a definite capacity, provided there is no change in the positions or electrical connections of other bodies near it. A charge Q given to it produces a proportional change of its potential. But the capacity of a conductor when not isolated is, of course, different from its capacity when isolated; for, when it is not isolated, inductive actions between it and other bodies can take place.

In any of these cases, by the capacity of an insulated conductor we mean *the magnitude of a charge given to it divided by the magnitude of the potential produced*, or, briefly, *charge per unit potential.*

394. The Charge on the Earth.—Electrical conditions over the surface of the earth and in the atmosphere vary widely with different states of the weather and also depend to some extent on the time of day and the season of the year. Neglecting these fluctuations, there is an average potential gradient of about 100 volts per meter or $\frac{1}{300}$ e.s.u. per cm. of elevation, the potential being higher at higher levels. This of course means that the body of the earth is surrounded by an electric field, the negative ends of the lines of force in the field being on the earth. Hence the surface of the earth, taken as a whole, carries a negative charge. If this were the only charge to be considered, the potential gradient near the earth's surface would vary only slightly for distances above the surface that are small compared with its radius. As a matter of fact, the gradient decreases rapidly with height. A few miles above the surface the gradient is practically zero, and so there is no electric field there and there are no lines of force. Hence all the lines have their positive ends somewhere in the atmosphere—on the water drops in

clouds or on the molecules of the air. The earth is therefore, in a sense, a condenser, one plate being the surface of the earth and the other the surrounding atmosphere.

The charge on the surface of the earth can be calculated from the average potential gradient near the surface and the area of the surface. If the whole charge on the surface is Q e.s.u. and the whole surface is A cm.², the number of lines of force per cm.² $4\pi Q/A$. This is therefore the field strength in e.s.u., which, as stated, is $\frac{1}{300}$ e.s.u. From this it is readily found that the charge on the surface is about 450,000 coulombs, which is, in reality, a small charge for a body of the size of the earth. Its presence does not in any way interfere with our using the potential of the earth as a practical zero of potential; for it is only with differences of potential that we are concerned in any practical problem.

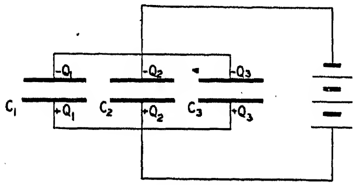


FIG. 280.

395. Condensers in Parallel and in Series.—Condensers are connected *in parallel* when they are arranged as in Fig. 280 for three condensers.

Let their capacities be C_1, C_2, C_3. When the system is charged by applying a potential difference V, the charges on the condensers are:

$$Q_1 = C_1V \qquad Q_2 = C_2V \qquad Q_3 = C_3V$$

Now if C is the capacity of a single condenser that would be equivalent to the group, the charge Q on it, or CV, would be equal to the sum of the charges on the separate condensers, or

$$\begin{aligned} Q &= Q_1 + Q_2 + Q_3 \\ CV &= C_1V + C_2V + C_3V \\ \therefore C &= C_1 + C_2 + C_3 \end{aligned}$$

The same method can evidently be extended to any number of condensers in parallel. *The total capacity of condensers connected in parallel is equal to the sum of their separate capacities.*

Condensers are connected *in series* when they are arranged as shown in Fig. 281 for three condensers. When the system is charged

by applying a potential difference V to the end plates, induction takes place and a quantity of electricity Q flows in each wire, W, connecting two plates. Hence the charges in the condensers are all equal to Q. The drops of potential in the successive condensers are Q/C_1, Q/C_2, Q/C_3 and the total drop is the sum of these. Now if C is the capacity of a single condenser that would be equivalent to the series, it would have the same charge Q, and the drop of potential in it would be Q/C. Equating this to the sum of the separate drops and dividing by Q, we get

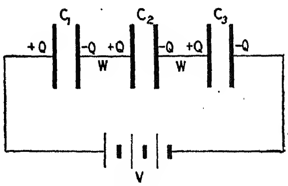


FIG. 281.

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}$$

In this case also we can extend the method to any number of condensers in series. *The reciprocal of the total capacity of condensers in series is equal to the sum of the reciprocals of their separate capacities.*

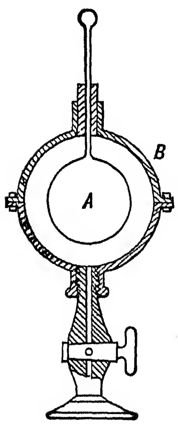


FIG. 282.

396. Measurement of Dielectric Constants.—The relation between the capacity of a condenser and the value of K for the dielectric suggests a much more accurate method for determining K for any medium than one founded on Coulomb's law (§368). The relation is, in fact, the basis of what is accepted as the best definition of K: *The dielectric constant of a substance* is *the ratio of the capacity of a condenser with that substance as dielectric to the capacity of the same condenser with a vacuum as dielectric.*

The effect of the dielectric on the capacity of a condenser was discovered by Faraday (1837). The following is an outline of the method he used. The apparatus consisted of two similar spherical condensers, like the one shown in Fig. 282, and a calibrated gold-leaf electroscope for measuring potential. The inner plate, A, was a brass sphere, the outer, B, a brass shell. B consisted of flanged hemispheres, which could be separated so that different dielectrics could be introduced, and there was a stop-cock for the admission of gases. The outer plates of both condensers were kept earthed.

Suppose one condenser, say S, had a solid or liquid dielectric, while the other, S', contained air. A charge, Q, was given to S, and the potential difference of its plates was found to be V. S was then, without being discharged, put in parallel with S', and the potential difference of the plates of each condenser became V'. Now let the capacity of each condenser with air as dielectric be C. What Faraday determined in his experiment was the dielectric constant of a substance relatively to air. Let us denote it by k. Then the capacity of S was kC and the capacity of S and S' in parallel was $kC + C$. Since the total charge was not changed by putting them in parallel

$$kCV = (kC + C)V'$$

Hence

$$k = \frac{V'}{V - V'}$$

By this means Faraday found k for various solid and liquid substances. He also tested various gases; but the instrumental means at his command were not sufficient to reveal the small differences in this respect between gases or between a gas and a vacuum. These have since been discovered by other methods. For example the dielectric constant of air, that is, K for air, is about 1.0006. To get K for any other substance from Faraday's value for k, we should multiply k for the substance by 1.0006. For all practical purposes the correction would be negligible. This is why we have, in several places, treated an electric field in air as being practically in a vacuum.

397. Energy of a Charged Condenser.—A charged condenser possesses energy equal to the work done in charging it or the work it can do in discharging. Evidence of this energy appears in the heat, light, and sound produced by the discharge. Large condensers used in high voltage work are dangerous to handle when charged.

To find the energy of a charged condenser, we may suppose that the condenser is initially uncharged and that the charging consists in transferring positive electricity in small instalments, q, from one plate to the other, leaving the latter with an equal negative charge. If the final potential difference of the plates is V and the charge Q, the mean potential difference against which Q is carried from one plate to the other is $\frac{1}{2}V$ and the whole work is $\frac{1}{2}QV$, which is

therefore the energy of the charged condenser. The process is represented graphically in Fig. 283, the area of the triangle representing the whole work of charging. Since $Q = CV$, where C is the capacity of the condenser, the energy can also be represented by other expressions, but it is perhaps best to remember it in the first simple form and transform it when necessary.

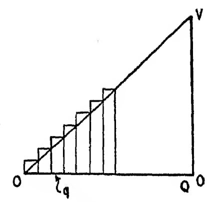


FIG. 283

In the expression $\frac{1}{2}QV$ or those derived from it, any consistent set of units may be employed. For example with electrostatic units of Q, V, and C, the energy is in ergs; when coulombs, volts and farads are used, it is in joules.

It is sometimes desirable to know the energy density, or energy per cm.^{3}, at a point in an electric field. For this purpose suppose the condenser we have been considering is a parallel plate condenser of area A and distance d between the plates (see Fig. 271). Its volume is Ad and therefore

$$\text{Energy density} = \frac{1}{2}\frac{QV}{Ad}$$

Now V/d is the field strength E between the plates. Also we have seen in §372 that $4\pi Q/K$ lines of force start from the charge Q. Hence the number of lines of force per cm.^{2} in the dielectric is $4\pi Q/KA$ and this is also equal to E. Thus Q/A in the expression for the energy density is equal to $KE/4\pi$. Hence

$$\text{Energy density} = \frac{KE^2}{8\pi}$$

While we have derived this result from a particular case, the only electrical quantities that appear in it are K and E, and it applies to any point in any electric field.

398. Piezo-electricity and Pyro-electricity.—When a quartz crystal is subjected to pressure, charges appear on the surface, positive on some faces and negative on others. The best effects are produced by a plate cut in a particular direction relatively to the optic axis of the crystal. Electricity produced in this way is called *piezo-electricity* (pressure electricity). Conversely, mechanical strains are set up in such a plate when it is electrified oppositely on the two faces by mounting it between two charged metallic plates. Many other non-isotropic crystals have similar properties.

This curious effect has found a number of practical applications, particularly as a means of converting mechanical vibrations into electrical vibrations and the converse.

If a crystal that has piezo-electric properties is heated or cooled, charges are produced on its surfaces. This is called *pyro-electricity* (heat electricity). The heating or cooling sets up mechanical strains in the crystal, and these produce the same electrical effect as the application of external forces in piezo-electricity.

399. Electrets.—The condition of a dielectric between the plates of a charged condenser is usually a temporary condition; it disappears when the field is discontinued, and the substance reverts to its ordinary condition. There are, however, certain substances, such as resin, carnauba wax, or beeswax, that show more permanent effects. If one of these is used as a dielectric of a charged condenser and is melted and allowed to solidify while in the field (5000 to 7000 volts per cm.), the plate, when removed from the condenser, is found to be positively charged on the face that was in contact with the negative plate of the condenser and negatively charged on the other face. It is called an *electret*. The charges on its surfaces are not just ordinary surface charges, like those on a glass rod when rubbed. They cannot be removed by contacts with other bodies, and, even if part of the surface is scraped off, the new surface is still charged. Such charges are called "bound" charges. They are the surface effects of the internal condition of the dielectric called polarization (§390).

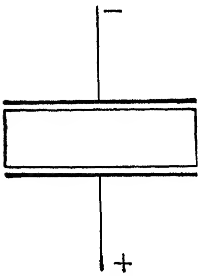


FIG. 284.

A peculiarity of these electrets is that their original polarity lasts, at best, for a few days only. It gradually fades out and is replaced by a reversed polarity that may last for years, if the electret is kept dry and shortcircuited. The reversal seems to be due to a piezo-electric effect that grows as the substance cools and internal strains are set up. The two different forms of polarity are probably due to different constituents in the complex mixture of which the substances that can form electrets consist.

400. How Large a Charge is the Electron?—After the discovery of the electron by J. J. Thomson in 1897, many attempts were made to determine the magnitude of this fundamental unit of charge. The difficulty of the measurement is shown by the fact that it was not until 1910 that R. A. Millikan, American physicist, succeeded, after several years of preparatory work, in completing a satisfactory determination. The method he used is called the "oil-drop" method because it depended on a study of the motion of charged oil drops in an electric field.

Very small drops of oil of radii between 10^{-5} and 10^{-4} cm. were sprayed by an atomizer A (Fig. 285) into the chamber B. They fell very slowly because of their small size and the viscous resistance of the air, just as the fine raindrops that form an ordinary cloud settle only very slowly. If the radius of a drop is a and its density ρ, its weight is $\frac{4}{3}\pi a^3 \rho g$ dynes. It is buoyed up by the weight of the air it displaces, which is $\frac{4}{3}\pi a^3 \sigma g$ dynes, where σ is the density of the air. Its motion is opposed by the viscous resistance of the air, which had been found by Stokes to be $6\pi a \eta v$ dynes, where η is the coefficient of viscosity of the air and v the velocity of the drop. Since the downward force on the drop is a constant force, while the resistance increases with the velocity, the drop reaches a steady velocity, say v, at which the two forces balance or

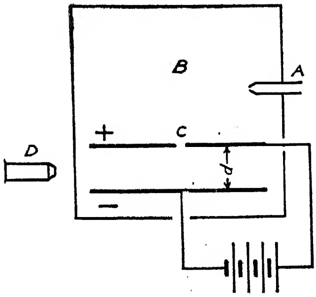


FIG. 285.

$$6\pi a \eta v = \tfrac{4}{3}\pi a^3(\rho - \sigma)g \tag{1}$$

Hence, if ρ, σ and η are known and v is observed, the radius a of the drop can be calculated. Evidently, if the force on the right-hand side is changed in any way, the velocity v will change in the same proportion.

In the experiment some of the drops became charged in the process of spraying. Occasionally one of these charged drops fell through a small hole C into the space between two metal plates that were connected to a battery. If the distance between the plates was d cm. and the potential difference V e.s.u., the field strength was $E = V/d$ e.s.u. The drop, when it was between the plates, was illuminated by a strong light (at right angles to the plane of Fig. 285) and was observed by the microscope D, which was provided with a scale. If the drop had initially a negative charge e e.s.u. and the upper plate was charged positively, there was an upward force Ee dynes exerted on the drop by the field. E could be adjusted until the drop remained stationary. The condition for this was that Ee should equal gravity minus buoyancy, or

$$Ee = \tfrac{4}{3}\pi a^3(\rho - \sigma)g \tag{2}$$

The air between the plates was then ionized by a radioactive substance or by X-rays Occasionally one of the charged particles struck the drop and adhered to it. This caused a sudden motion or change of motion of the drop, and its speed was found by means of the microscope. Usually the motion of the drop could be followed while it received a great many successive charges, some negative and some positive. The change of speed after an additional charge was proportional to the additional charge and could be taken as a measure of it. It was thus found that all the increments of charge were either equal to a certain minimum or were small integral multiples (2, 3, 4, 5, .) of the minimum, and this minimum could be nothing but a single electron.

To complete the calculation of the electronic charge e it was now necessary to know a in equation (2). For this purpose the battery was disconnected and the drop allowed to fall Its speed of fall was measured and substituted in equation (1) This gave a, and substitution in (2) then gave e.

As this brief outline gives little idea of the extent of such a piece of work, it may be added that Millikan studied the motions of several hundred drops of different oils in different gases at different pressures.

The value of e now accepted as accurate to about one tenth of one per cent is

$$e = 4.803 \times 10^{-10} \text{ e s.u}$$

MAGNETISM

401. Magnets.—The science of magnetism appears to have begun with the observation that an iron ore, Fe_3O_4, has the power to attract small fragments of iron to itself. This ore, well known to the ancient Greeks and Romans, was plentiful in the district of Magnesia near the Aegean coast. For that reason it came to be known as magnetite. Substances having the property of attracting small pieces of iron are called *magnets* and the property is called *magnetism*. This name is also applied to the science which has evolved from the study of phenomena associated with magnets.

As early as 1100 B.C. the Chinese were familiar with the fact that a piece of magnetite, suspended in such a manner that it can swing freely in a horizontal plane, will set itself pointing approximately in a north and south direction and form a primitive mari-

ner's compass. The use of magnetite as a compass does not appear to have been known in Europe until some 2000 years later. Then the name lodestone (leading stone) came to be used in English for magnetite.

In discussing the simpler phenomena associated with magnetism, a magnetized bar of iron will serve our purpose better than a piece of lodestone, which now is of purely historical interest. A bar of iron can be artificially magnetized by stroking it several times with one end of a piece of lodestone or, very much more efficiently, by electrical means. The electrical method is discussed in detail later (§439).

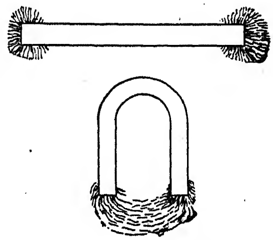


FIG. 286.

402. Magnetic Poles.—If a magnet, natural or artificial, is dipped into iron filings, it is found that there are certain regions where the filings cling in tufts to the magnet and others where none are held. The regions where the tufts are greatest and hence the attraction for the filings greatest are called the *magnetic poles* of the magnet. It is found that a magnet always has at least two poles, and, if there are more than two, they occur in pairs. The usual type of bar magnet has two poles, situated near the ends of the bar, the middle being a neutral region. The line joining the two magnetic poles is called the *magnetic axis* of the magnet.

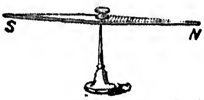


FIG. 287.

Suppose a magnet is made in the form of Fig. 287, in which the magnetic axis coincides with the geometrical axis, and is then suspended on a frictionless bearing, so that it can rotate freely in a horizontal plane. This magnet or compass needle will then come to rest pointing approximately north and south, if no other magnets are in the vicinity. But the magnet does not tend to move as a whole in either direction, as may be shown by placing it on a cork and floating it on water (Fig. 288). When it has turned so as to point about north and south, it comes to rest, the tendency of one pole to move toward the north being exactly counterbalanced by the tendency of the other to move in the opposite direction. Hence the two poles are equally strong, though a definition of pole strength may be left until later (§405).

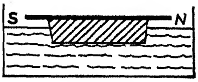


FIG. 288.

The pole which points northward is called, in English speaking countries, the *north*, N, or north-seeking pole and the other pole the *south*, S, or south-seeking pole. (In France the names are interchanged.) The fact that a freely suspended magnet always sets itself with the same pole pointing north indicates that there is an essential difference between the two poles. This may be shown by bringing the north pole of a magnet, first near to the north pole of the compass needle, and then near to the south pole of the needle. It is found that the two north poles repel each other, but the north pole of the magnet attracts the south pole of the compass needle. Similarly it may be shown that the two south poles repel each other. Hence, not only are the two poles of a magnet different from one another, but also *like-named poles repel each other and unlike-named poles attract each other.*

If a magnet is magnetized so that it has more than two poles, an equal number of N and S poles exist, and there are no other kinds of magnetic poles than those called north-seeking and south-seeking poles. A simple method for determining whether a bar of iron is magnetized or not is to bring it near a magnet. If it is possible to get repulsion between the bar of iron and one of the poles of the magnet, the bar of iron is magnetized. If only attraction occurs, the iron is unmagnetized.

403. Magnetization.—When an iron bar is magnetized, there is no increase in its weight; so the process of magnetization is, presumably, a rearranging of the atoms or molecules of the iron. Physical changes, such as changes in length and volume, do occur on magnetization, and these also point to some rearrangement within the body of the iron. If a bar magnet is cut in two, two complete magnets are always produced; no further subdivision will produce a N pole without an accompanying S pole. Moreover the two poles of each small magnet produced are always of equal strength.

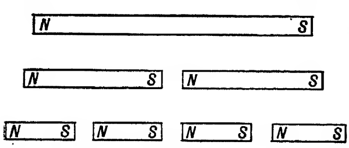


Fig. 289.

These facts seem to show that the poles of a magnet are merely the surface effects of an internal condition. The fact that cutting up a magnet always yields complete magnets suggests that a magnet of ordinary size consists of numerous very small magnets, lined up in the direction of the axis of the magnet. We may suppose that,

before the bar is magnetized as a whole, the elementary magnets already exist in the iron, but are turned, pretty much at random, in all directions, though likely to be arranged in local groups (Fig. 290). Magnetization, then, would consist in making the elementary magnets wheel around into a common direction, the end of the bar where the elementary N poles are outward becoming the N pole of the magnet, and the other end its S pole.

The poles of a magnetized bar are not, in general, exactly at the ends of the bar. They are usually about $\frac{5}{6}$ of the length of the bar apart, but the actual distance depends on several factors, such as

Fig. 290.

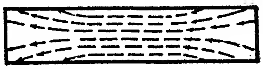


Fig. 291.

the length of the bar relatively to its cross-section and the strength to which it is magnetized. These things can also be explained by the assumption of permanently magnetized particles. Near an end of the bar the repulsion between like poles deflects some of them toward a side surface (Fig. 291).

There is no reason to doubt the general correctness of this explanation of some of the elementary phenomena of the subject; but it is to be noticed that it does not tell us what the permanently magnetized particles are or why there should be such things. This will be considered later (§449).

404. Induced Magnetization.—We have seen in Electrostatics that one charged body can be used to charge a second body without contact, namely by induction. In a somewhat similar way a magnet can be used to magnetize an unmagnetized body, such as iron, without stroking or contact but by a process called *magnetic induction* or induction of magnetization. For example, if a piece of soft iron or soft steel, such as a nail, is brought near a pole of a magnet (Fig. 292), it is attracted and it also acquires the property of attracting other nails, so that a series of them can be supported, each becoming a magnet by induction.

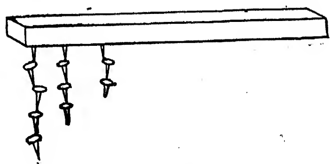


Fig. 292.

A second resemblance, but only a partial resemblance, between magnetic induction and electrostatic induction can readily be

verified. If the nails are of very soft iron and if the upper one is held in the fingers and the magnet is removed, the chain of nails usually drops apart, which means that the magnetization of each was only temporary, just as in the case of a conductor electrified by induction. But if the nails be of harder iron, the chain will show some tendency to hold together, indicating that part, at least, of the magnetization is permanent. Very hard iron will retain nearly all the magnetization induced in it, and it is by using the property that the magnets we shall have occasion to refer to are made.

But there is one respect in which electrostatic induction and magnetic induction are wholly different. When a conductor is electrified by induction, something flows, namely electrons. If the conductor is cut in two while under induction, the two halves have opposite charges. In magnetic induction there is no flow of anything, though in earlier times magnetism was supposed to be a fluid. If an iron rod under induction is cut in two, two complete magnets result, which shows that the magnetization is a condition existing everywhere in the rod.

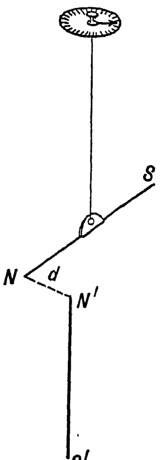


FIG. 293.

405. Coulomb's Law of Force in Magnetism.—The establishment in 1785 of the law of force between magnetic poles by Coulomb (to some extent anticipated by others) laid the basis for an exact treatment of what had been a rather vague subject. Coulomb's work in this investigation was more difficult than in the similar case of electric charges (§368) because a N pole can never be wholly isolated from the associated S pole. To overcome this difficulty, Coulomb used very long thin steel magnets, so that the force between two poles on different magnets could be tested while the influences of the other poles were small.

One magnet, NS, was suspended by a silk thread to serve as a torsion balance (Fig. 293), and a pole of the other, $N'S'$, was brought near a pole of NS. If they were like poles repulsion took place, and N moved away from N'. It was then brought back to its initial position by turning the torsion head and thereby twisting the thread, the total twist being proportional to the force of repulsion of N and N'. In this way the forces at different distances were compared. Coulomb also assured himself, by a variety of pre-

liminary experiments, that any magnet has a definite pole strength; for example two magnets used side by side in the position of $N'S'$ exert on N a force equal to the sum of the forces they exert separately. We can now state Coulomb's law: *Two magnetic poles of strengths m_1 and m_2 placed d cm apart in a vacuum exert on each other a force of magnitude given by*

$$F = \frac{m_1 m_2}{d^2}$$

One immediate result of Coulomb's work was that it became possible to define a unit of pole strength and thereafter state pole strengths of magnets in such a way that they could be used in calculations. *A unit magnetic pole is one which, placed one centimeter distant from an equal and similar pole in a vacuum, repels it with a force of one dyne.*

The force between two poles when they are in some medium is not the same as when they are the same distance apart in a vacuum. This is taken care of, in the formula for Coulomb's law, by introducing a factor of $1/\mu$ and writing the law in the form

$$F = \frac{m_1 m_2}{\mu d^2}$$

The property of the medium denoted by μ is called the *permeability* of the medium. Although we first meet it in Coulomb's law, it has much greater practical importance in another relation (§441); and its use in that connection will explain the choice of the name permeability and why in Coulomb's law μ is introduced in the denominator, not the numerator, of the right hand side. The case is similar to that of K in electrostatics (§368), and here also it will be better to defer the exact definition of μ and an account of how it is measured until later. At present it will be sufficient to say that the values of μ for air and other gases differ so little from 1 that we shall usually omit μ in Coulomb's law and formulae derived from it.

406. Magnetic Fields.—Just as we have found it convenient to consider the space around a gravitating body, like the earth, as a gravitational field and the space around electric charges as an electric field, so too it is useful to refer to the space surrounding magnets as a magnetic field. The test of whether a part of space is a magnetic field is whether a magnet placed there would tend to set itself in any particular direction. If it does, the direction in which

its N pole is urged is taken as the *direction* of the field. The magnitude or *intensity* of the field, usually denoted by H, *is measured by the force exerted by the field on unit pole.* Although we are here referring to a single pole as though it could be isolated, which we know to be impossible, we shall see that the definition does not involve us in any difficulty, and it usually helps us to simplify things by using forces instead of moments of forces or torques.

To be able to state the intensity of a field numerically we need a unit of field intensity. Here, following the same method as in Electrostatics, we define unit field intensity as *the intensity of a field that would exert unit force on a unit pole.* This unit is called the *oersted* (in honor of the Dane who discovered the magnetic effects of currents in 1820). The intensity of the magnetic field H at a distance of r cms. in a vacuum from a pole of strength m is m/r^2 oersteds. If the pole is immersed in a medium of permeability μ the field H is $m/\mu r^2$ oersteds. The quantity μH is called the *flux density* or *magnetic induction,* B, at the point and is measured in *gausses.* A detailed discussion of the relation of B and H is given later (§440). From the definition of the oersted it follows that a pole of strength m in c.g.s. units, when in a field of H oersteds, is acted on by a force of mH dynes. In the preceding, both definition and formula are limited to cases in which the magnet does not itself disturb the field appreciably. We can now see why it is that the experiment of the magnet floating on water (§402) in a uniform magnetic field allows us to conclude that the poles of a magnet are of equal strength.

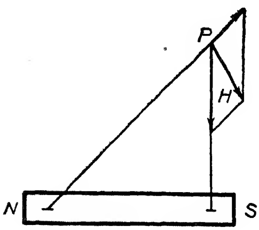


FIG. 294.

Since magnetic field intensity, like force, is a vector quantity, in adding field intensities we must use the vector method of addition, namely the parallelogram method or something equivalent to it. Thus at a point in the field of a single magnet the field intensity is the vector sum of the intensities due to the two poles (Fig. 294). If, however, the point is on the axis of the magnet, the addition is made by means of ordinary algebra.

407. Lines of Force in a Magnetic Field.—The method of lines of force for representing a field of force graphically is used for a magnetic field in the same way as for an electrostatic field. At any point on a magnetic line of force the direction of the line is the

direction of the field; it is the direction in which a N pole would tend to move and opposite to that in which a S pole would tend to move. A small compass needle placed at a point on a line of force in a uniform field will point along the line of force and will have no tendency to move sidewise. But if the field is not uniform, the lines of force are curved, and the forces on the two poles will not be exactly in opposite directions. The needle, while pointing along the line of force through its center, will then have some tendency to move sidewise into the stronger part of the field.

A very easy way of actually plotting the lines of force in some simple cases is by the use of iron filings. For example, for the case

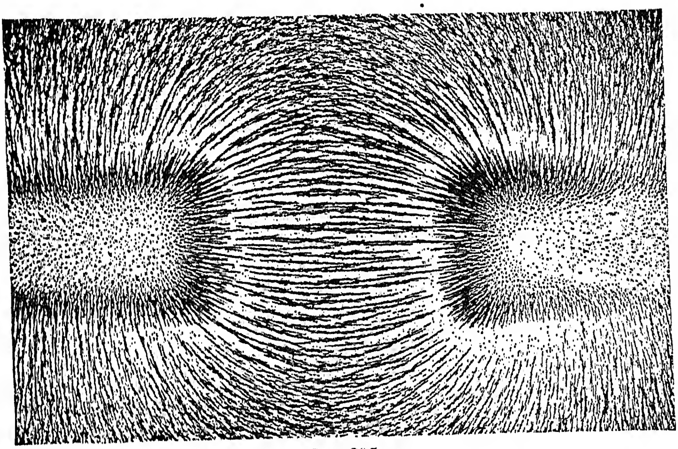


Fig. 295.

of the field of two unlike poles (Fig. 295), a plate of glass is placed on the magnets and iron filings are sprinkled on it. A tap on the plate will make the filings jump into positions along the lines of force. Each filing becomes magnetized by the field by induction, and they line up along a line of force, like a series of small compass needles. Here, as in an electrostatic field, the lines of force present the appearance of a tension along a line and pressure across it. It is evident from the arrangement of the filings that the poles are not exactly at the ends of the magnets. The effect of another tap or two illustrates the statement just made with regard to the forces on a small magnet in a non-uniform field, for at each tap the filings move sidewise into a stronger part of the field. Fig. 296 shows the field between two N poles, and here we note again the appearance

of a lateral pressure. The field of a complete magnet is shown in the same way in Fig. 297.

There are, however, important differences between the lines of force in a magnetic field and those in an electrostatic field. A

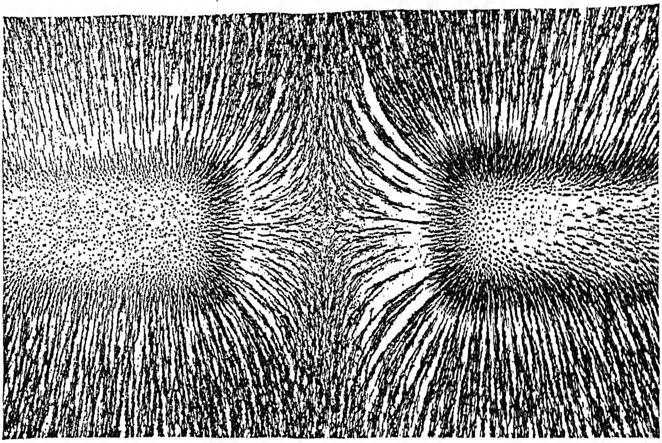


FIG. 296.

conductor with a positive charge can be isolated, that is, it can be removed from physical contact with any other body (though there is somewhere an equal and opposite charge); but a N pole

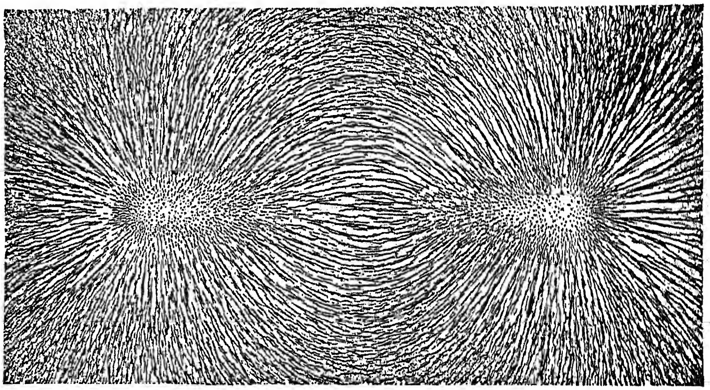


FIG. 297.

cannot be isolated, it is always accompanied by a S pole on the same body. There is no electric field inside the charged conductor (if it does not enclose a charge), but there is a magnetic field

inside a magnet. The lines of force in an electrostatic field end on charges. The magnetic lines of the field of a magnet enter the S pole and continue as lines of induction, as they are called (§440) through the iron to the N pole where they emerge, so that each line is a completely closed loop. (Fig. 298).

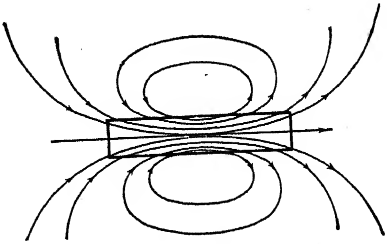


FIG. 298.

Another difference between the electric field and the magnetic field is that electric charges, if free to move, start off along electric lines of force and form a current. Nothing similar occurs in the magnetic field; there is nothing that can be called magnetic conduction, corresponding to electric conduction.

408. Magnetic Lines of Force and Field Strength.—We have, so far, used lines of force in a qualitative way, merely to show the *direction* of the field. We can, however, use them as a measure of the *strength* of the field. The method is like that used for an electrostatic field (§372). In both cases it is founded on the inverse square law; and, as the reasoning is the same in both cases, it will be sufficient to state the results for a magnetic field.

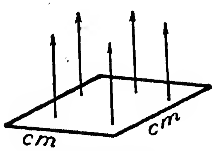


FIG. 299.

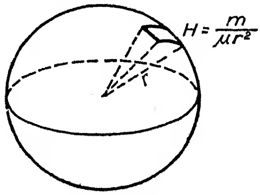


FIG. 300.

Suppose that at any point in the magnetic field lines of force are drawn such that *the number of lines of force per cm.² at right angles to the field is numerically equal to the field strength there;* so that, for example, if the field strength is 5 oersteds (Fig. 299) there are 5 lines of force per cm.². We then find that the number of such lines of force issuing from a pole of strength m in a medium of permeability μ is $4\pi m/\mu$ (Fig. 300). The field strength at any other point in the field (not within a magnet) is numerically equal to the number of lines of force per unit area at right angles to the field.

409. Torque on a Magnet in a Magnetic Field.—When a magnet is in a uniform magnetic field, such as the earth's field, with its axis at an angle θ to the direction of the field (Fig. 301), there are equal and opposite forces acting on the poles and these form a couple. If the strength of the field is H, the forces on the poles $+m$ and $-m$ are $+mH$ and $-mH$, and if l is the distance between the poles and L the moment of the couple

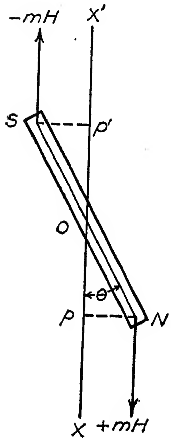


FIG. 301.

$$L = mH \times (\text{arm of couple})$$
$$= mH \times (NP + SP')$$
$$= mH \times (ON + OS) \sin \theta$$
$$= mH \times l \sin \theta$$

Here the product ml is a constant for the magnet. It is called the magnetic moment of the magnet and is denoted by M. *The magnetic moment of a magnet is the product of its pole strength and the distance between the poles,* or

$$M = ml$$

Hence

$$L = MH \sin \theta$$

If $\theta = 0$, $L = 0$, and if $\theta = 90°$, $L = MH$ (Fig. 302). If the magnetic moment of the magnet is 1 and the magnet is perpendicular to a uniform field of intensity H, L and H are numerically equal. This relation might have been taken at the outset as the definition of field strength, so as to avoid the difficulty about an "isolated" pole; but the two definitions would be equivalent.

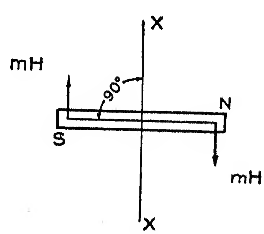


FIG. 302.

When a magnet is in a field that is not uniform, the forces acting on the two poles may not be equal and opposite. They are, however, equivalent to a couple plus a force, and the force tends to make the magnet move as a whole in some direction.

410. Period of Oscillation of a Suspended Magnet.—Field strengths can be compared by a method that resembles the use of a pendulum for testing differences in the earth's gravitational field. Let us suppose that we wish to compare the strengths of the earth's magnetic field at two stations. A small magnet is suspended hori-

zontally by a fine thread so that it is free to vibrate horizontally in the earth's field. When it is turned so that it makes an angle θ with the direction of the earth's magnetic field, there is a couple, $MH \sin \theta$, acting on it and tending to turn it into the direction of the earth's field. Hence, when released, it oscillates. If the angle θ is small, $\sin \theta$ may be taken as equal to θ (in radian measurement) so that the restoring couple exerted by the earth's field is $MH\theta$. Hence if we put $L = -MH\theta$ and denote the rotational inertia of the magnet by I and the angular acceleration by α

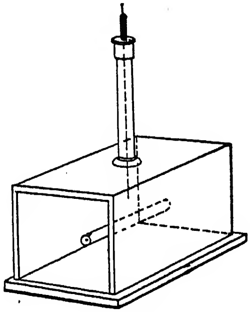


FIG. 303.

$$L = I\alpha = -MH\theta$$

so that

$$\alpha = -\frac{MH}{I}\theta.$$

Thus the angular acceleration is proportional to the angular displacement, and from this it follows (§123) that the oscillations are angular harmonic oscillations. Hence if the period of oscillation is T

$$T = 2\pi\sqrt{\frac{I}{MH}}$$

The period T is therefore inversely proportional to the square root of H. If the period is T_1 in a field of strength H_1 and T_2 in a field of strength H_2

$$\frac{H_1}{H_2} = \frac{T_2^2}{T_1^2}$$

If the frequencies of the oscillations in the two fields are n_1 and n_2, $n_1 = 1/T_1$ and $n_2 = 1/T_2$. Hence

$$\frac{H_1}{H_2} = \frac{n_1^2}{n_2^2}$$

Since n_1 and n_2 can be determined with great accuracy by timing a large number of vibrations, this is a very accurate method for finding H_1/H_2. While we have referred to it as a method of studying the earth's field, it is evidently applicable to other fields, such as a horizontal section of the field of a magnet or a group of magnets.

411. Field Strength at Points in the Field of a Magnet.—There are certain cases in which the field strength at a point in the field of a magnet NS can readily be calculated from Coulomb's law if we are given the pole strength m and the distance between the poles, say $2l$ (to simplify expressions).

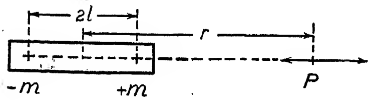


FIG. 304.

Consider first a point P on the axis, extended through N. The whole field strength H at P is the algebraic sum of two components, due to the two poles, and from Coulomb's law

$$H = \frac{m}{(r-l)^2} - \frac{m}{(r+l)^2} = \frac{4rlm}{(r^2-l^2)^2} = \frac{2Mr}{(r^2-l^2)^2}$$

There is, however, a practical difficulty in using this result. We have seen earlier (§403) that there is always some uncertainty in knowing what to take for the distance, $2l$, between the poles of a magnet. The effect on H of this uncertainty as regards l diminishes as r is increased. Let us suppose that r is so great or l so small (or both) that l/r is very small and l^2/r^2 wholly negligible compared with 1. We may then write:

$$H = \frac{2Mr}{r^4\left(1 - \frac{l^2}{r^2}\right)^2} = \frac{2M}{r^3} \tag{1}$$

Next consider a point P on the line through the center O of the magnet and perpendicular to the axis (Fig. 305). Here again the

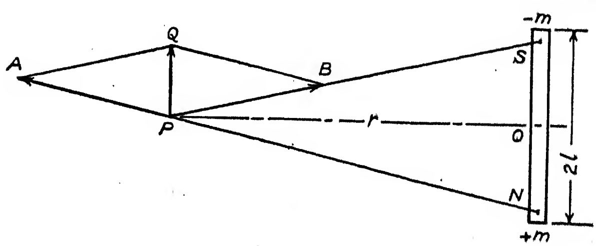


FIG. 305.

resultant field strength, say H', is the sum of components due to the two poles, but they are not now in the same line. One is in the direction NP, the other in the direction PS. But they are of the same magnitude, m/d^2, where d is the distance of P from a pole, and we can represent them by equal lines PA and PB. Their resultant

will then be represented by the diagonal PQ of the parallelogram $PAQB$. We now note that the triangles PAQ and NPS are similar, so that

$$\frac{PQ}{PA} = \frac{NS}{NP}$$

In this PQ and PA represent H' and m/d^2 respectively; $NS = 2l$ and $NP = d = (r^2 + l^2)^{\frac{1}{2}}$. Substituting and remembering that $2lm = M$, we get

$$H' = \frac{M}{(r^2 + l^2)^{\frac{3}{2}}}$$

Here there is the same difficulty as regards l as in the first case. But if we again suppose l/r small and l^2/r^2 negligible, we get

$$H' = \frac{M}{r^3} \tag{2}$$

In both of the preceding cases there would also be uncertainty about the value of M, if it had to be found from measurements of m and $2l$. There is, however, no practical difficulty, for, as we shall see presently, M can be found by experiment and much more accurately than m or $2l$ could be determined.

Comparing equations (1) and (2), we see that $H = 2H'$. This result, really deduced from Coulomb's law, would not be correct unless that law were correct. We shall see presently how it can be used for a rigorous test of the law (§412).

412. Resultant of Two Fields.—In considering the field of a single magnet we have, for simplicity, considered it by itself as if there were no other magnets with their fields anywhere near. As a matter of fact we can hardly exclude the presence of one other field, namely that of the earth. The simplest case of the addition or superposition of two fields to form a resultant field is when both fields are uniform fields, or sufficiently nearly uniform in a limited region, and they are at right angles.

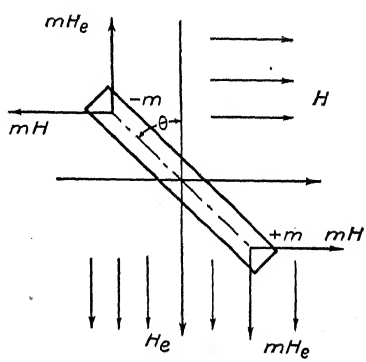


Fig. 306.

For definiteness we shall suppose that one of the two fields is the horizontal part of the earth's field. Let us denote its field strength by H_e and the field strength of the other field by H. If the second field were not present, a magnet suspended so as to turn freely would come to rest in the direction of H_e. The presence of the other field deflects it and it comes to rest in the direction of the resultant of the two fields. Let the angle by which it is turned away from the direction of the earth's field be θ. It is then in equilibrium under the influence of two equal and opposite couples. If its magnetic moment is M, the moments of the couples (§409) are $MH_e \sin \theta$ and $MH \cos \theta$. Hence

$$MH_e \sin \theta = MH \cos \theta$$

and therefore

$$\tan \theta = \frac{H}{H_e}.$$

or

$$H = H_e \tan \theta$$

Thus, if H_e is constant, H is proportional to $\tan \theta$. This is often called briefly "the tangent law."

We can now complete the statement at the end of §411 as regards testing Coulomb's law. If the same magnet be placed at equal (large) distances from a compass needle, first "end on" and then "broadside on" (§411), $\tan \theta$ should be twice $\tan \theta'$, where θ and θ' are the angles of deflection of the compass needle with the magnet in the two positions. This was essentially the test applied by Gauss in 1833. (In reality he went further and took account of l^2/r^2 but not of l^4/r^4, etc.)

413. Measurement of the Earth's Field.—Accurate knowledge of the earth's magnetic field is of great practical importance and special means have been devised for obtaining it. One of the most important quantities to be found is the horizontal intensity. We shall now denote it by H (without the subscript e which we used earlier for clearness). It might be thought that it could be found by timing the vibrations of a small suspended magnet (§410) but it will be noticed that it is tied up with M, which must be found by some other experiment. Or we might try to use the tangent law and compare the field of the earth with the field of a magnet, but this would again require M for the magnet. It will therefore

be seen that two different experiments involving H and M must be made, and then both H and M can be calculated from the results.

The timing of the vibrations of a suspended magnet at the place where it is desired to determine H is chosen as one of the two experiments, for this can be done with a high degree of accuracy. This gives us a first equation in H and M as two unknowns:

$$T = 2\pi\sqrt{\frac{I}{MH}} \tag{1}$$

In this the rotational inertia I can be calculated from the weight and dimensions of the magnet.

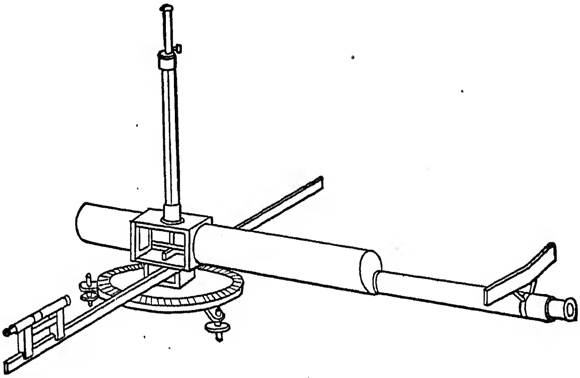


FIG. 307.

The second experiment is the using of the same magnet to deflect a compass needle as considered in §411. If the end-on position is used, we get, by applying the tangent law, a second equation in H and M:

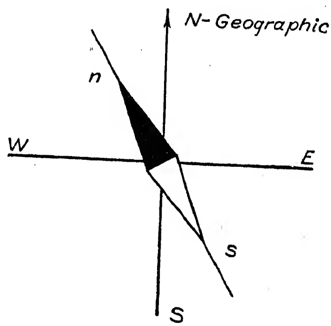
$$\frac{2M}{r^3} = H \tan \theta \tag{2}$$

From these two equations we can eliminate M and get an expression for H in terms of T, I, r, θ, all of which are measurable. We can also find M at the same time, and this is the method referred to in §411 for finding M by experiment, instead of trying to calculate it from the pole strength and length of a magnet.

The second of the two experiments is performed by the use of a magnetometer (Fig. 307). It consists of a small magnet, carrying a mirror and suspended at the center of a graduated bar that can be

turned exactly at right angles to the earth's field H. A scale and telescope are used for reading deflections of the magnet by reflection from the mirror. Of course, in very accurate work, much more elaborate apparatus is used.

414. Declination and Dip.—These are the other two quantities which, with H, are known as the "elements" of the earth's magnetic field. A compass needle does not point geographically north and south (Fig. 308). The angle by which it deviates therefrom is called the *declination* of the magnetic field. It varies widely over the surface of the earth as indicated in Fig. 309. Lines passing through places of equal declination are called *isogonic lines*. It will be



Declination West

FIG. 308.

noticed how extremely irregular these lines are compared with the corresponding geographic lines, namely the great circles of equal longitude. The isogonic line of zero declination is called the agonic line. It runs fairly centrally from north to south down through both North and South America. At points east of it the declination is to the west of the geographic north, while at points west of it the declination is easterly. The discovery that there is such a thing as the declination is usually said to have been made by Columbus in his first voyage west (1492), though possibly what he discovered was that the declination, which was easterly in the Mediterranean at that time, had changed to a westerly one before he reached America.

A compass needle stays in a horizontal plane because it is on a vertical axis. A magnetic needle on a horizontal axis through its center of gravity acts differently. (Fig. 310.) Over the northern

hemisphere (roughly speaking) it points N pole downward, and over the southern it points N pole upward. The angle it makes with the horizontal at any place is called the *dip* or *inclination* of the earth's field at the place. Lines of equal dip are called *isoclinic lines;* the line of no dip or aclinic line corresponds roughly to the Equator.

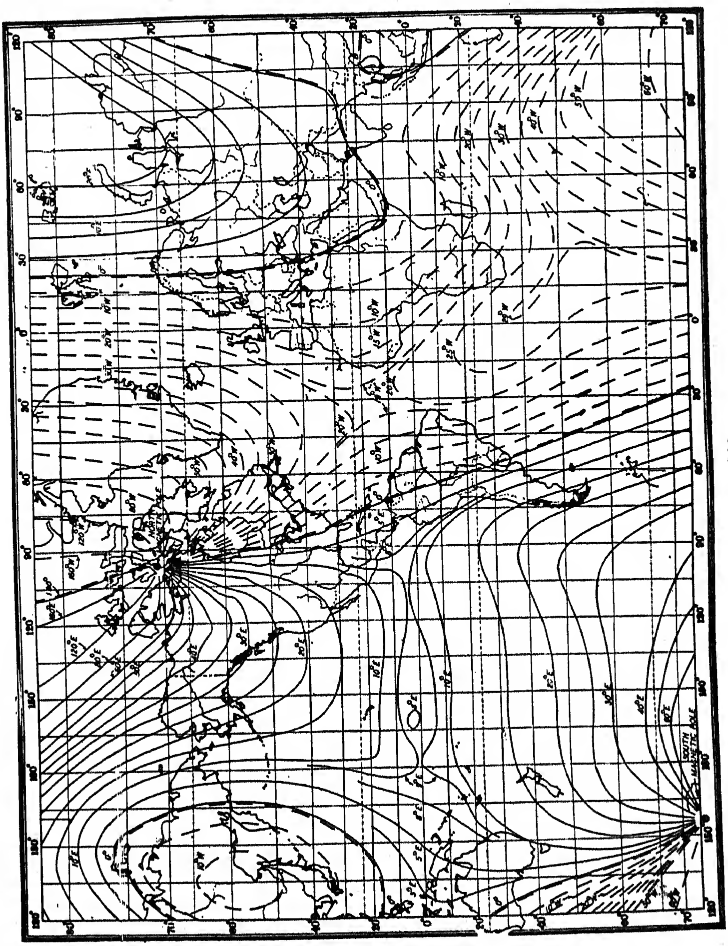


FIG. 309.

The fact that there is a dip shows that the real direction of the earth's magnetic field at a point is inclined to the horizontal, the horizontal intensity which we denote by H being only the horizontal component of the real intensity T, the other component, V, being

vertical. The relation between T, H, V and the dip ϕ is shown in Fig. 311. At Washington at present $\phi = 71.2°$, $T = 0.5712$, $H = 0.1835$, $V = 0.5409$ oersted.

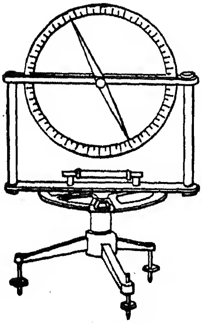


FIG. 310.

415. Variations of the Earth's Magnetic Condition.—We have been speaking of the earth's magnetic "constants" as if they were constant. As a matter of fact they are changing all the time and in a variety of ways. These *variations* are given different names corresponding to the different ways in which they show themselves and the rates at which they take place. The most fundamental one is the so-called *secular* (age-long) variation. It is evidenced chiefly by changes in the declination and dip at any place, for these have been recorded for several hundred years. For example the change of declination and dip at London since 1540 are shown in a graphical way in Fig. 312. Apparently these changes would complete a cycle in about 500 years. With these changes there is, of course, a change in the position of what we called the magnetic north pole of the earth, which is not a point but a rather small area over which a dipping needle points about vertically downward. Apparently it rotates slowly about the geographical pole in a period of about 500 years. Or one may say that the magnetic axis of the earth (again a rather indefinite term) rotates about the geographical axis, requiring probably about 500 years for a complete rotation. Of course with these changes go changes in isogonic and isoclinic lines. For example the agonic line (no compass "variation") passed through Washington, D. C., at about 1800 A.D. and the declination at Washington has now become 6° W. What trouble such changes have caused in using old surveys can be imagined.

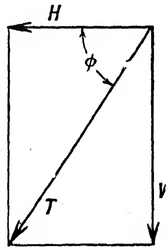


FIG. 311.

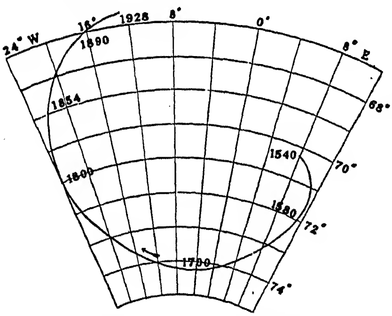


FIG. 312.—Secular variation of declination at London.

There is also a variation of the magnetic elements having a period of a year, called the *annual* variation, and a *daily* variation.

These, however, are small (fractions of a degree) compared with the secular variation, and they are undoubtedly due to external bodies, sun, moon, and planets. The preceding are all periodic, but there are irregular and comparatively violent disturbances, called *magnetic storms*, and as they occur when there are violent disturbances on the surface of the sun, as shown by sunspots, they are certainly an effect of external origin.

From this brief account it is seen that the earth is a permanent magnet, though what made it so and what keeps it so we do not know; but on this permanent magnetization is superimposed an induced magnetization due to external bodies.

ELECTRIC CURRENTS

416. Three Kinds of Currents.—In considering the properties of electricity at rest it was necessary sometimes to assume that electricity can move in bodies called conductors (§367). Electric charges in motion constitute electric currents, and it is convenient to distinguish three kinds, though they may exist in combinations.

(1) If the current consists of a motion of charged particles through a body or along the surface of a body, the current is called a *conduction* current. In a metal only free electrons or conduction electrons move forming a current. In a liquid or gas, both positively and negatively charged particles can flow. The discovery of conduction currents was made by Galvani (1791) and Volta (1800).

(2) When charges are moved by mechanical means they constitute a *convection* current. An example is the stream of charges in the Van de Graaff generator carried on the belt into the large sphere. That such a stream of charges has the same properties as a conduction current was first proved by Rowland (§426).

(3) A varying electric field causes slight shiftings of the positions of electrons in atoms and, while these are taking place, they constitute a *displacement* current. Even when there is no matter present, the changing electric field has an effect that is also called a displacement current. It was the recognition of displacement currents that led Maxwell (§523) to the electromagnetic theory of light and later gave rise to radio waves.

417. Conduction Currents.—Suppose that in the arrangement of Fig. 235 the plates between which there is a potential difference are replaced by a wire. Free electrons then flow in the wire, moving from the end at low potential to the end at high potential.

Their motion constitutes an *electric current,* which shows its presence by heating the wire and by other effects to be considered later. In spite of the fact that the current is a stream of electrons from low potential to high potential (Fig. 313), we take the opposite direction, from high potential to low potential, as the direction of the current; it is the direction in which positive electricity would move if it were free to do so. This use of the term *direction of a current* was adopted long before anything was known about electrons, and to change it now would produce confusion.

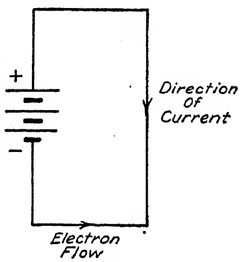


FIG. 313.

We shall see later that there are methods by which the magnitude of the stream of electrons can be measured in terms of some unit of flow, and instruments are constructed to read it. The practical unit of current, the *ampere,* will be defined later. Meantime it will be sufficient to say that if Q coulombs flow through each cross-section of the wire in t seconds, the flow per second Q/t is the current in amperes. Hence, briefly

$$\text{coulombs/seconds} = \text{amperes}$$
$$\text{coulombs} = \text{amperes} \times \text{seconds}$$

418. Motion of Electrons in Conduction.—A current in a wire is a motion or drift of electrons in the direction opposite to the applied e.m.f. It is not an accelerated drift, because the electrons collide with the atoms (or molecules) and rebound in random directions. This interference with the forward motion of the electrons constitutes the electrical resistance of the conductor.

Suppose a conductor is of uniform cross-section A cm.^{2} and that it contains n free, or conduction, electrons per cm.^{3}, each with a charge e. Let their average forward speed be v. The electrons that pass through a cross-section in a second are those in a length v of the conductor, and their number is therefore Anv, and the charge they carry through the cross-section is $Anev$. Hence if the current is i

$$i = Anev$$

From this v can be calculated, if i, A, n, e are known. For example, suppose i is a current of 10 amperes in a copper wire of one square mm. or 0.01 cm.^{2} cross-section. There is reason to believe that the number of free electrons per cm.^{3} is about equal to the number of atoms per cm.^{3} and for copper this is known to be[1] 8.28×10^{22}. The charge of an electron, e, is 1.60×10^{-19} coulomb (§400). Substituting, we get $v = 0.076$ cm./sec. Thus the speed of drift is a very small fraction of the average velocity of the electrons in random directions, which is known to be very large.

[1] 63.6 gms. of copper contain 6.06×10^{23} atoms; the mass of 1 c.c. of copper is 8.7 gms. and it contains 8.28×10^{22} atoms.

419. Electromotive Force.—When electrons move through a wire they collide with atoms in the wire and give up energy to the atoms. The increased energy of motion of the atoms is observed as a rise of temperature. Thus heat is produced, and there must be some source of energy in the electrical circuit to supply the energy of the current that is transformed into heat. The source of energy may be a voltaic cell, an electromagnetic generator or a thermocouple. The cell transforms chemical energy into the energy of the current, the generator transforms mechanical energy, and the thermocouple transforms heat energy. In each case the agent is regarded as a source of an *electromotive force* that gives rise to the current. The term electromotive force or e.m.f. is not wholly satisfactory, but it has now become firmly embedded in the language. It is not a force in the mechanical sense, and it is used with somewhat different meanings in different connections. Its meaning in each case is really defined by the way in which it is measured and is indicated by some qualifying phrase.

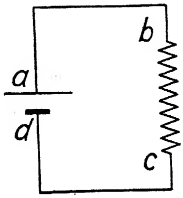


FIG. 314.

Now consider the arrangement in Fig. 314. Electricity travels in a closed circuit *abcd*. In the wire there is a fall of potential, say V. There is also a fall of potential, say V', in the liquid of the cell. The sum of these, $V + V'$, is the total *fall* of potential in the circuit. There is, however, no net fall the whole way around

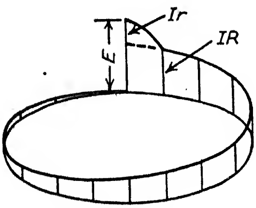


FIG. 315*a*.

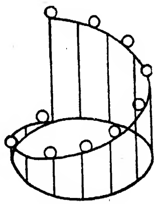


FIG. 315*b*.

the circuit; for at the contacts of different substances there are abrupt changes of potential. We do not need to consider these more closely at present. It is sufficient to say that, though there may actually be a fall of potential at one contact, the total effect at all contacts is a *rise* of potential equal in magnitude to $V + V'$. Remembering that potential difference means work done on unit charge taken from one potential to another, we can now define *the*

e.m.f. around the circuit. It is: *the work done on, or work done by, unit quantity of electricity taken once around the circuit.* It is also equal to the algebraic sum of the abrupt changes or jumps of potential at plates and liquid; this is called *the e.m.f. of the cell.* Also it is equal to the sum of the falls of potential in wire and liquid. The different rises and falls of potential in the circuit are as represented in Fig. 135*a*, which suggests the analogy of a roller-coaster (Fig. 315*b*), with its changes of potential energy of the car. The analogy, though not perfect, is instructive (§422).

The e.m.f. E of a cell, or the algebraic sum of the abrupt changes of potential at the contacts of different substances, can be found by connecting the cell to an electrometer by wires of the same material. The cell is not then in a closed circuit and there is no fall of potential in the liquid.

While we have defined the term electromotive force as it applies to a complete circuit, the term can, without any confusion, be applied to a part of a circuit. For example, the potential difference that exists between the ends of the wire *bc* in Fig. 314 may be regarded as the cause of the current in the wire and is also called the electromotive force acting on the wire. There are, however, cases in which a current in a conductor is not due simply to a potential difference (§483), and we must then distinguish between e.m.f. and p.d.

From the definitions of potential difference and electromotive force, both are of the nature of *work per unit charge.* Hence the units of potential difference that we have already defined (§375) serve also as units of electromotive force.

420. Ohm's Law, Resistance.—The relation between an electromotive force and the current that it produces is a matter of great practical importance; and it is fortunate that, at least for solid and liquid conductors, to which we shall confine attention at present, the relation is a very simple one: *Current is proportional to electromotive force.* When an e.m.f. or potential difference is applied to a wire, the current increases exactly in proportion to the increase in e.m.f. *provided* the physical condition of the wire (temperature etc.) is unchanged. This was first discovered by G. S. Ohm, German physicist (1789–1854) in 1827 and is known as *Ohm's law* for a conductor. Besides its simplicity, the law is remarkable for the range of currents over which it holds. It is accurate for the feeblest currents that have been measured, and it holds (for silver) up to currents of at least 10,000 amperes per mm.^{2}.

The ratio of e.m.f. to current for a conductor is called the *resistance* of the conductor. Ohm's law is equivalent to the statement that the resistance of a conductor (solid or liquid) in a constant physical condition is constant. If we denote e.m.f. by E, current by i and resistance by r, Ohm's law states that

$$\frac{E}{i} = r \qquad \text{or} \qquad E = ir$$

But the proviso that the physical condition of the conductor must be constant is, as we shall see, a very important one.

The practical unit of resistance is called the *ohm*; one volt applied to a conductor of one ohm resistance produces a current of one ampere. For precision in legal matters the ohm is defined as the resistance, at 0°C. of a uniform column of mercury that is 106.300 cm. long and weighs 14.4521 gm.

421. Ohm's Law for a Circuit.—From Ohm's law for the separate conductors (solids or liquids) in a circuit in which there is a battery,

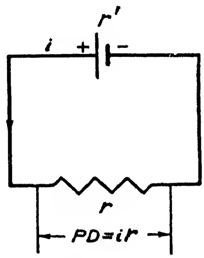


FIG. 316.

we can derive a more general form of the law, applicable to a complete circuit. Let the e.m.f. of the battery be E. It is equal to the sum of the potential drops, V and V', in the external resistance r and the internal resistance r' of the battery. If the current is i amperes, $V = ir$ and $V' = ir'$. Hence

$$E = i(r + r')$$

and if we denote the whole resistance of the circuit by R,

$$i = \frac{E}{r + r'} = \frac{E}{R} \qquad \text{or} \qquad E = Ri$$

This is the form of Ohm's law for the whole circuit. As in the case of any broad fundamental generalization, the real proof of its accuracy is that all deductions from it are found to be verified.

422. Energy Expended in Maintaining a Current.—There are different ways in which a flow of electricity in a circuit can be maintained and we shall consider them later. For the present let us assume that the generator is a battery. Electricity is raised in potential by the battery and flows back to lower potential through the conductors in the circuit. Work has to be done at the expense of the chemical energy of the battery to raise the potential of the electricity; and the current does work, which, by the conservation of energy, is equal to the work done by the battery. We may compare this with an analogous case in Mechanics. When the car of a roller-coaster (Fig. 315*b*) is raised, it has potential energy that it received from the motor. The amount of this potential energy, per unit mass of the car, corresponds to the potential, *i.e.* potential energy per unit charge, to which the electricity is raised by the battery. As the car slides down the incline, the initial potential energy is converted into heat.

When a steady current i flows for t seconds, the quantity of electricity that passes through each cross-section of the circuit is it, and this is also the quantity that is raised through a potential difference E by the battery. Hence if the work done by the battery or the work done by the current is W

$$W = Eit$$

If E and i are in electromagnetic units, W is in ergs. If E is in volts and i in amperes, W is in joules (§375). Rate of working or work per second, W/t, is called *power* and is denoted by P. Hence

$$P = Ei$$

When electromagnetic units are used P is in ergs per second. In practical units P is in joules per second or *watts*. Hence, briefly,

$$\text{Watts} = \text{Volts} \times \text{amperes}$$

423. Work Done by a Current.—The expression for the energy supplied by the battery or the work done by the current can be changed by means of Ohm's law, $E = ir$. We thus get in either electromagnetic or practical units

$$W = i^2rt$$

In this W is the work done in all parts of the circuit where there are resistances; it includes both the external parts of the circuit and the liquid in the battery. We can separate it into parts corresponding

to the separate resistances. Thus if the external circuit contains a wire of resistance r the work done by the current in the wire is given by the above equation, E being the p.d. at the ends of the wire. In this case the energy expended in the wire produces heat, that is, added energy of motion of molecules due to the impulses given to the electrons by the e.m.f. To find the heat produced in calories from the work in joules, we may use practical units and divide W by the mechanical equivalent of heat, which is 4.18 joules per calorie. Hence, if H is the heat produced in time t,

$$H = i^2rt/4.18 = 0.239i^2rt \text{ calories}$$

in which i is in amperes and r in ohms. This is called *Joule's law.*

424. Magnetic Effects of Currents. Oersted's Discovery.—Until 1820 no one suspected that there was any connection between Electricity and Magnetism. They were wholly different sciences. Even now there has never been any discovery of a relation between charges of electricity and magnetic poles *if both are at rest.* But a charge in motion is a current. Oersted of Copenhagen noticed in 1820 that a current in a wire deflected a compass needle that was near it (Fig. 317) and that reversing the current reversed the deflection. As we would say now *an electric current is always accompanied by a magnetic field.* This was a very fruitful discovery.

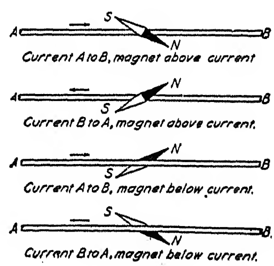


FIG. 317.

A magnetic field is represented by magnetic lines of force. In the case of a current in a long straight wire the magnetic lines of force are circles surrounding the wire with their centers on the wire, as in Fig. 318 where the current is supposed to be downward perpendicular to the paper. A small compass needle, placed in the field, would tend to set itself tangentially to the circle through its center. A convenient way of remembering the direction of these magnetic lines is the "corkscrew rule," also called the right-hand screw rule or screwdriver rule. (Fig. 319). If the thrust of a corkscrew is in the direction of the current, its twist is in the direction of the lines of force. The

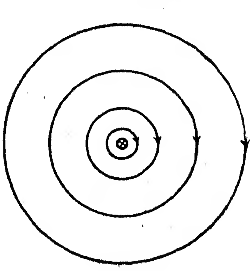


FIG. 318.

nature of the field can also be shown by the device of iron filings on a horizontal glass plate surrounding a vertical current (Fig. 320).

Oersted's discovery was the magnetic field of a conduction current, but we shall see later that convection currents and displacement currents also have magnetic fields.

425. Two Illustrations of Magnetic Fields of Currents.—When a wire that carries a current is bent into a circular form, the magnetic lines of force pass through the area bounded by the circle, entering at one face of the circle and going out through the other (Fig. 321). These

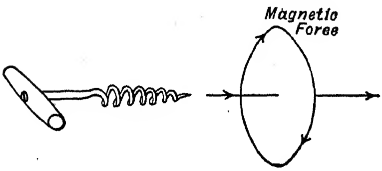


Fig. 319.

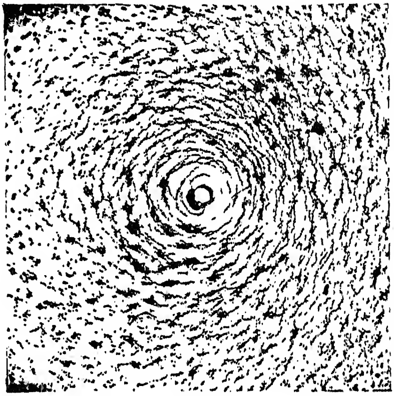


Fig. 320.

magnetic lines are closed loops, but not circles with their centers on the wire. Their direction can be found by the corkscrew rule.

By winding insulated wire closely on a cylinder (Fig. 322) a helix or *solenoid* is formed. The lines of force of each turn are

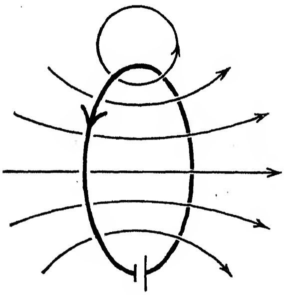


Fig. 321.

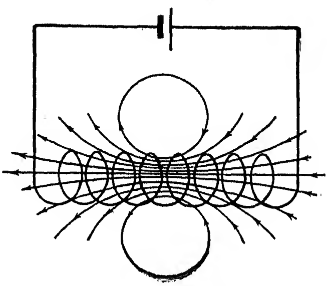


Fig. 322.

continued through the other turns and produce a uniform field except near the ends, where the lines of force "leak out" from the solenoid.

A solenoid carrying a current acts like a magnet. The direction of its magnetic axis is the direction of the magnetic lines through it.

Its N pole is where these lines leave the solenoid, its S pole where they enter. If suspended on an axis with freedom to turn in a horizontal plane, it acts like a compass needle. It is a striking indication that magnetism must at basis be electrical in its nature.

426. Convection Currents.—The brief account we have just given of what we know about the nature of an electric current in a wire summarizes the results obtained by many physicists in different countries in the last hundred years. An exceptionally important link in the chain of evidence was supplied by Rowland in 1876.

If a conduction current in a wire is a stream of minute charged particles (now called electrons), an electric charge on a rapidly moving non-conductor should produce a magnetic field. Rowland's test, in reality a very difficult one, consisted essentially in rotating a charged vulcanite disk at a very high speed and showing that it deflected a magnetic needle. Great care had to be taken to shield the needle from jars, air currents and electrostatic effects, and provision had to be made for detecting minute deflections of the needle (by a reflection of light method). The results showed that "electricity produces the same magnetic effect in the case of convection as of conduction."

Several other physicists have repeated the test and always with the same result: a *convection current*, as it is called, is equivalent magnetically to a conduction current.

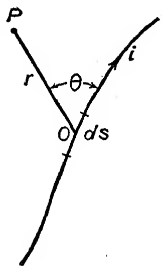


FIG. 323.

427. Ampere's Law for the Magnetic Field of a Current.—Soon after Oersted's discovery Ampere found, as the result of a large number of experiments, a very useful formula for calculating the magnetic field strength H at any point P in the magnetic field of a current (Fig. 323). He supposed the whole length s of the circuit to be divided into small parts or elements, any one of which we may denote by the symbol ds, and then combined the contributions of each element, say dH, to get the whole field strength H. If P is at a distance r from ds, and if θ is the angle OP makes with ds, Ampere's formula is

$$dH = k\frac{ids \sin \theta}{r^2}$$

In this k is some constant the magnitude of which will depend on the units used in the formula. For a given value of r, dH is a

maximum when $\theta = 90°$ and it is zero when $\theta = 0$. The direction of dH at P is related to the direction of i in ds by the corkscrew rule. In the figure dH is up.

This is a very general formula applicable to a circuit of any form. We shall, however, use it only for a circuit that lies in one plane. If P is also in the plane, all the contributions dH are perpendicular to the plane of the current and H is found by adding them by ordinary algebra.

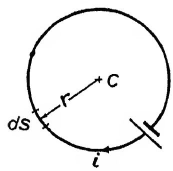


FIG. 324.

The simplest application of Ampere's formula is to find the magnetic field strength H at the center C of a circular coil of radius r that carries a current i. In this case the short parts ds of s are equidistant from C, $\sin\,\theta = 1$, and $s = 2\pi r$. Summing for both sides of Ampere's formula we get

$$H = k\frac{si}{r^2} = k\frac{2\pi ri}{r^2} = k\frac{2\pi i}{r}$$

428. Electromagnetic Units and Practical Units.—In stating Ampere's law and deriving from it a formula for the field strength at the center of a circular circuit the choice of unit of current was left open. The unit adopted will determine the value of the constant of proportionality k. It is a special convenience to have k unity, so that it may be omitted. As regards the H in the formulae we have already adopted a unit of it, namely, the oersted (§406). The unit of current that will provide for $k = 1$ is called the *electromagnetic unit of current.* It is defined as: *the current in a circular circuit of* 1 *cm. radius that produces a magnetic field strength of* 2π *oersteds at the center.*

With this definition of the unit of current the formula for the field strength at the center of a circular loop of radius a becomes

$$H = \frac{2\pi i}{a}$$

and for a circular coil of n turns

$$H = \frac{2\pi ni}{a}$$

We can now also define the electromagnetic unit of quantity of electricity. It is the unit that satisfies the equation

$$Q = it \qquad \text{or} \qquad i = \frac{Q}{t}$$

when i is in e.m.u. and t in seconds, and from this a formal definition can readily be constructed.

As stated earlier, the practical unit of current, the *ampere*, is $\frac{1}{10}$ of an e.m.u. of current, and the practical unit of quantity, the *coulomb*, is $\frac{1}{10}$ of an e.m.u. of quantity.

Two other electromagnetic units can now be defined by means of equations with which we have already become familar. For the electromagnetic unit of electromotive force the defining equation is

$$W = EQ \qquad \text{or} \qquad E = \frac{W}{Q}$$

where W is in ergs and Q in e.m.u. of charge. For the e.m.u. of resistance the defining equation is Ohm's law

$$i = \frac{E}{r} \qquad \text{or} \qquad r = \frac{E}{i}$$

From these equations formal definitions of the units of E and r can readily be constructed.

In the practical system of units the unit e.m.f., the volt, is equal to 10^8 e.m.u., and the unit of resistance, the ohm, is 10^9 e.m.u. With these units Ohm's law holds for practical units. The powers of 10 used enable us to take the mechanical equivalent of heat as simply 4.18 (§423).

429. Field of Current in Long Straight Wire. As another application of Ampere's law let us find the magnetic field strength at a point P near a long straight wire that forms part of a circuit in which there is a current i. If we suppose that the rest of the circuit is at a great distance away compared with the distance d of P from the wire LM we can calculate H at P by applying Ampere's law to the wire alone. The result of summing up the contributions from all the elements of the wire gives for H

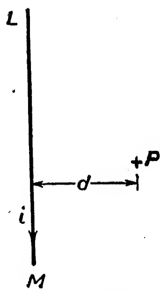


FIG. 325.

$$H = \frac{2i}{d} \text{ oersteds}$$

In this we have assumed that the current i is in e.m.u. As it takes 10 amperes to make 1 e.m.u., if i is in amperes the right-hand side must be divided by 10.

The only satisfactory way of performing the summation is by means of the integral calculus. To integrate Ampere's expression for dH we must first express it in terms of a single variable, say the angle θ. This is easily done if we note that

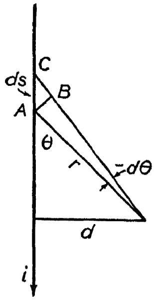


FIG. 326.

$$ds \sin \theta = AB = -r d\theta \qquad \text{and} \qquad r \sin \theta = d$$

We thus get

$$dH = -i \sin \theta \frac{d\theta}{d}$$

and on integrating this from $\theta = \pi$ to $\theta = 0$ we get $2i/d$.

A direct proof that H varies inversely as d is given by a simple experiment (due to Maxwell). AOB is a vertical wire that carries a current i downward, the circuit being completed at a large distance. NS is a magnet that rests on a horizontal disk, which can rotate about the wire. If the magnetic field strengths at the poles, due to the current, are H_1 and H_2, the forces exerted on the poles are mH_1 and $-mH_2$; and, if the distances of the poles from AOB are d_1 and d_2, the moments of the forces are mH_1d_1 and $-mH_2d_2$. When a current is switched on, the disk does not move. Hence

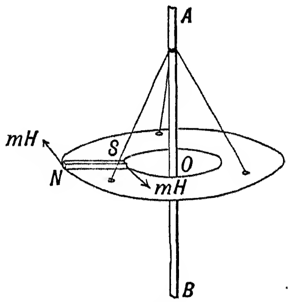


FIG. 327.

$$mH_1d_1 - mH_2d_2 = 0$$

and therefore

$$\frac{H_1}{H_2} = \frac{d_2}{d_1}$$

or the field strength varies inversely as the distance from the wire.

430. Work in Path Around a Current.—The magnetic lines of a current in a long straight wire are circles with their centers on the wire. Imagine a unit N pole to be carried once around the wire following a circular line of force of radius r. If the current is i e.m.u. the force acting on the pole at each point is $2i/r$, and the length of the path is $2\pi r$. Hence the work is $4\pi i$ ergs, a result that is independent of the radius of the path. If, however, i is in amperes the work will be $4\pi i/10$ ergs.

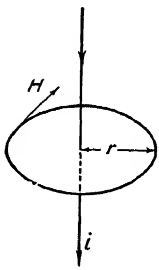


FIG. 328.

This result can readily be extended to the case in which the path followed is not a circle but is any closed path. For any short

step along the path can be replaced by three steps: one along a circle as before; a second parallel to the current; and a third radial. The work in the second and third steps would be zero since there is no force in either direction, and, as the work in a circle is independent of the size of the circle, it is readily seen that the whole work is $4\pi i$ as before.

While we have established this $4\pi i$-law, as it is called, only for a special case, it can be shown from Ampere's law that it is perfectly general and quite independent of the form of the circuit in which the current i flows. This extremely simple and very useful relation, deducible from Ampere's law, is sometimes called the "circuital form of Ampere's law."

431. Field in a Solenoid.—A solenoid consists of a large number of turns of wire wound on a cylinder. When a current flows in it, the magnetic field in each turn is much stronger than it would be for a single circle, for the neighboring turns contribute. Magnetic lines of force run through the solenoid and are completed in loops outside.

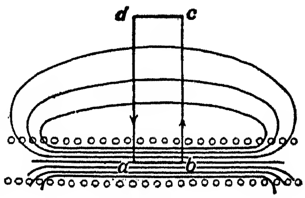


FIG. 329.

Consider a closed path *abcd*, *ab* being parallel to the axis, *bc* and *ad* perpendicular to it, and *cd* so far away that the field along it is negligible. If a unit pole be taken around this closed path, work is done only along *ab* and $n \cdot ab$ turns are encircled, where n is the number of turns per cm. Hence (§430)

$$W = H \cdot ab = 4\pi i \cdot n \cdot ab$$

and therefore

$$H = 4\pi ni$$

H is in oersteds if i is in e.m.u. If i is in amperes the right-hand side must be divided by 10.

The result we have obtained is for points in the solenoid and not near its ends, and it is the same at all points in a cross-section. From the plot of the magnetic lines of force (Fig. 322) it is seen that many of them "leak out" before reaching the ends of the solenoid. In fact, at an end H is only $2\pi ni$, as can be shown by deriving a formula from Ampere's law in its first form (the mathematical work is not difficult but it is lengthy and must be omitted here).

A toroidal solenoid (Fig. 330) is ring-shaped and has no ends. Though there may be some leakage of lines of force between turns, it is slight if the winding is close. Consider now a path of radius r inside the solenoid. When a unit pole is taken around it, the work done is $H \cdot 2\pi r$, and it is also (by the $4\pi i$ law) $4\pi i \cdot n \cdot 2\pi r$. Hence $H = 4\pi ni$. A closed path in the position of l_1 or l_2 would not enclose (link with) any turns of the solenoid, so that the work would be zero. Hence the solenoid has no external field.

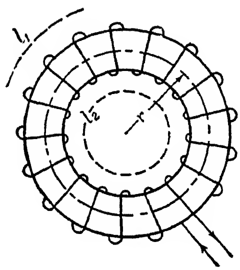


Fig. 330.

If for either form of solenoid we write the expression for H in terms of i expressed in amperes, then

$$H = \frac{4\pi ni}{10} = \frac{4\pi Ni}{10l}$$

where N is the total number of turns and l the length of the solenoid. The product Ni is called the *ampere-turns* of the solenoid.

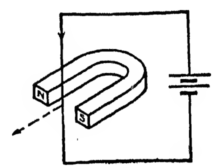


Fig. 331.

432. Force on a Current in a Magnetic Field. Since a current exerts a force on a magnetic pole, it would naturally be expected that a magnetic field would, by the principle of action and reaction, exert a force on a wire that carries a current. This converse effect was discovered by Ampere. It is readily demonstrated by the action of a horseshoe magnet on a straight wire that carries a current (Fig. 331). The wire moves across the lines of force of the magnet; and, if the current is reversed, the movement of the wire is also reversed. This can readily be explained by the magnetic lines of force of the field (Fig. 332). The lines of force of the magnet are from the N pole to the S pole, those of the current are circles with their centers on the wire. On one side the two fields are in the same direction, giving a strengthened field with lines of force closer together; on the other side they oppose and the field is weakened.

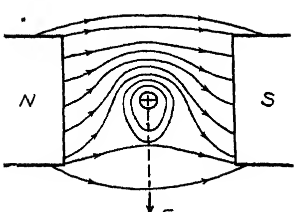


Fig. 332.

The force exerted on the wire may be regarded as due to the tension in the curved lines of force.

433. Formula for the Force on a Current.—By applying the principle of action and reaction, we should be able to derive from the formula for Ampere's law (§427) an expression for the force exerted on a current by a magnetic field. Let us suppose that there is a N pole of pole strength m at P (Fig. 333). The force exerted on it by the magnetic field dH at P, due to the current i in ds, is mdH. The force exerted by the pole m on ds must be mdH in the opposite direction. Let us denote this force on ds by dF. Then

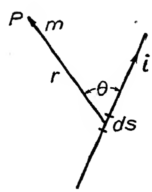


FIG. 333.

$$dF = mdH = \frac{m}{r^2} ids \sin \theta$$

If H is the magnetic field strength at ds due to m at P and μ is the permeability of the medium (§406)

$$H = \frac{m}{\mu r^2} \quad \text{and} \quad \frac{m}{r^2} = \mu H$$

Hence

$$dF = \mu H ids \sin \theta = Bids \sin \theta$$

We have here replaced the product μH for a magnetic field by a single letter B (§406). It denotes the *flux density* of the field and is, as we shall see later (§440) a quantity of great importance.

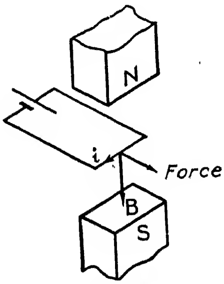


FIG. 334.

As regards the direction of the force dF exerted on ds, we note that in Fig. 333 the force exerted by the current in ds on the pole m at P is upward in the figure. Hence the force exerted by m on ds is downward. Notice also that H and B at ds are directed from P to ds.

The force exerted on a length l of a straight wire carrying a current i in a uniform magnetic field is obtained by summing the two sides of the last equation. For the important case in which θ is 90°, we get

$$F = Bil, \text{ (i being in e.m.u.)}$$

This is illustrated by Fig. 334 in which the directions of F, B, and i are indicated. The relations of these directions are summarized in what is called the *left hand rule* or *motor rule:* Suppose the left

hand to be held with the forefinger, the center finger and the thumb mutually at right angles, and so that *the forefinger is in the direction of the flux B and the center finger in the direction of the current i; then the thumb is in the direction of the force F or the motion it would produce.*

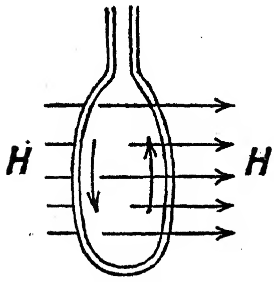


FIG. 335.

434. Couple on a Coil Carrying a Current.—If a coil (all in one plane) that carries a current is suspended in a uniform magnetic field, it will tend to turn in some definite direction. It is equivalent to a magnet with an axis perpendicular to the plane of the coil, and it tends to turn so that this axis is in the direction of the field; that is, so that the plane of the coil is perpendicular to the field. In this position the magnetic flux through it, due to its own current, is in the same direction as the field flux. Thus *it tends to turn so that the total flux through it is a maximum.* This is called Maxwell's rule (Fig. 335).

If a rectangular coil $PQRS$, with sides vertical and horizontal (Fig. 336a) and of lengths p and q respectively, is at an angle θ to a magnetic field of flux density B (Fig. 336b) and carries a current of

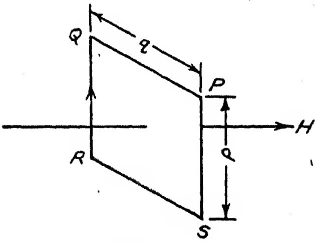


FIG. 336a.

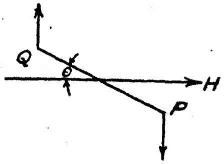


FIG. 336b.

i e.m.u., the magnitude of the force acting on each vertical side is Bip and these forces then form a couple, the arm of which is $q \cos \theta$. Hence, if we denote the moment of the couple by L,

$$\begin{aligned} L &= Bipq \cos \theta \\ &= BAi \cos \theta \end{aligned}$$

where A is the area of the coil. This is a maximum, BAi, when the coil is parallel to the field ($\theta = 0$) and is zero when the coil is perpendicular to the field. As it is only the area A of the coil that

enters the equation, the result should be the same whatever the shape of the coil. This is confirmed by more complete analysis.

435. Tangent Galvanometer.—Galvanometers are instruments for measuring or detecting a current by means of its magnetic effects. The tangent galvanometer consists of a large coil in a vertical plane, mounted so that it can be turned about a vertical axis, and a small compass needle with a scale at the center of the coil.

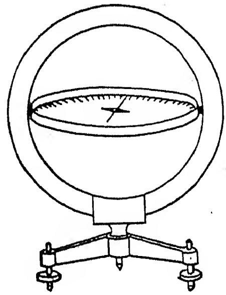


Fig. 337a.

The coil is first turned so that, when there is no current in it, the needle and the coil are in the same vertical plane. The scale reading is then zero. When there is a current of i e.m.u. in the coil the needle is deflected. It comes to rest in the position in which the moment of the couple exerted on it by the magnetic field, say H_c, of the current in the coil balances the moment of the couple exerted on it by the earth's field H_e. Denote its deflection by θ. Since the two magnetic fields are at right angles (Fig. 337*b*)

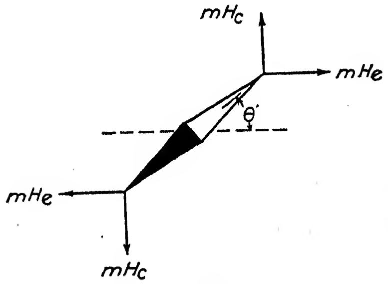


Fig. 337*b*.

$$H_c = H_e \tan \theta$$

If the coil has n turns of radius r, $H_c = 2\pi ni/r$. Substituting, we get

$$i = \frac{H_e r}{2\pi n} \tan \theta$$

Here i is in e.m.u. If it is in amperes, it must be divided by 10 before substitution in the equation. The result is that the same equation applies, i being in amperes, if the right-hand side is multiplied by 10.

A tangent galvanometer can be used for comparing currents by comparing the values of tan θ; but its chief use has been in getting the actual magnitude of a current. For this purpose the value of H_e at the place must be known or found by measurement (§413).

436. Moving-coil Galvanometer.—A galvanometer of this type measures or detects a current by means of its magnetic effect, but, as compared with the tangent galvanometer, the parts played by

coil and magnet are reversed, the coil rotating while the magnet is fixed.

A small coil that carries the current is suspended by a very thin metal strip (of phosphor-bronze or even gold) and hangs between the poles of a very strong U-shaped magnet (Fig. 338). When there is a current in the coil, it rotates until arrested by the torsion of the suspension.

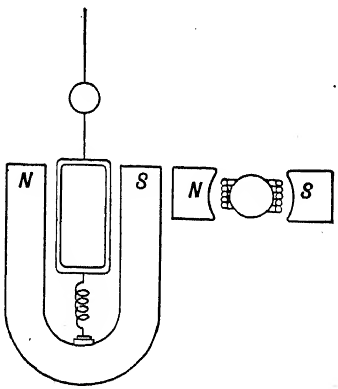


FIG. 338.

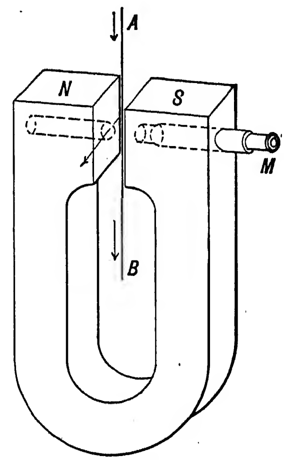


FIG. 339.

While the moving-coil galvanometer can be used for comparing or even measuring very small currents, its chief use is for detecting a current (§453). Its sensitiveness for this purpose is increased by attaching a small mirror to the coil and reading with a telescope the reflection of a distant scale. For a given current the deflection is greater the stronger the magnetic field. To strengthen the field there is a small stationary iron cylinder within the coil, and the pole pieces are curved so as to concentrate the field on the coil. The motion of the coil is "damped" by induced currents (§491) in a closed loop of wire carried by the coil. In this way the index can be made "dead-beat" (no vibration).

The ordinary moving-coil galvanometer, used for electrical measurements, gives a scale reading of 1 mm at a scale distance of 1 meter for 10^{-8} ampere; and 10^{-10} ampere has been reached in the best instruments of this type. Still greater sensitiveness, 10^{-12}, has been reached by the Einthoven galvanometer (Fig. 339).

It is founded on the principle of Fig. 331. A very fine wire between the poles of a U-shaped magnet carries the current and is observed by a microscope.

437. Ballistic Galvanometer.—It is sometimes necessary to measure a quantity of electricity, for example a condenser discharge. There is a similarity between this problem and that of finding the momentum of a bullet using a ballistic pendulum. For this reason a galvanometer for measuring a quantity of electricity is called a *ballistic galvanometer.* In calculating the momentum of a bullet from the swing of a ballistic pendulum into which it has been fired, it must be assumed that the bullet comes to rest in the pendulum before the pendulum has moved appreciably. Similarly, a ballistic galvanometer should have a somewhat long period, so that the discharge through it is complete before there is any appreciable deflection. The swing θ of the galvanometer will then be proportional to the charge Q or

$$Q = k\theta$$

where k is called the ballistic constant of the galvanometer.

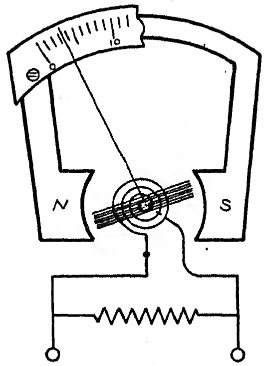


Fig. 340.

An ordinary moving-coil galvanometer can be used ballistically, but one requirement is that the damping should be small; so the damping loop is removed. Theoretically it would be better to have no damping at all, but some is unavoidable. To correct for it the ratio of successive free swings, θ_1/θ_2, θ_2/θ_1, etc. on right and left is observed. This is called the damping constant λ. Then, if θ is the first swing when discharge takes place, its corrected value is $\lambda\theta$.

438. Ammeter. Voltmeter. Wattmeter.—An ammeter measures a current, a voltmeter measures a potential difference. Those in common use are based on the principle of the moving-coil galvanometer, but the construction is much more rugged. The coils are supported on agate pivots and the control torque is exerted by a coiled spring. Both instruments are highly damped and therefore dead-beat.

The only essential difference between an ammeter and a voltmeter (of moving-coil type) is as regards their resistances. In an

ammeter (Fig. 340) the coil is shunted, the resistance of the shunt being about $\frac{1}{100}$ ohm. Thus the ammeter does not change appreciably the current to be measured. In the voltmeter there is a very high resistance, of the order of 15,000 ohms, in series with the coil, so that the instrument does not lower appreciably the potential difference to be measured. A wattmeter may be regarded as an ammeter and a voltmeter combined (Fig. 342). There is no permanent magnet, the magnetic field in which the coil moves being provided by the line current by means of the coils A and B. The coil C is connected across the power line through a very high resistance R, so that the current in C is proportional to the voltage. Thus the deflection is proportional to both current and voltage and is therefore a measure of their product or watts.

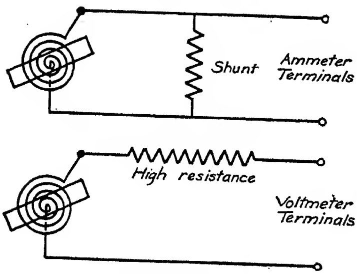


FIG. 341.

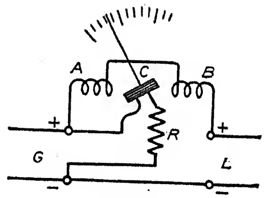


FIG. 342.

MAGNETIC INDUCTION

439. Magnetization of Iron by a Current.—We have already seen (§425) that a solenoid carrying a current produces a magnetic field somewhat similar to that of a permanent magnet. The strength of the magnetic field produced by the solenoid may be tested by bringing it near a suspended magnetic needle. If a rod of soft iron is inserted into the solenoid the deflection of the needle is greatly increased. The rod, coil and current form an *electromagnet.*

The magnetic field produced by the current in the solenoid causes the elementary magnets of the iron bar to aline themselves with the field of the solenoid. Thus the iron becomes magnetized and produces a magnetic field in the same direction as that of the solenoid. The resultant field is the sum of that due to the solenoid and that due to the magnetism induced in the iron bar. To be able to study and describe the condition of *magnetized material,* such as the iron of the rod in the magnetizing coil, we need a term that does not depend on the dimensions of the specimen, just as in the ordinary study of matter we need such terms as the density and elasticity of the material. One term used for the purpose is *inten-*

sity of magnetization. It is defined as *magnetic moment per unit volume* and is denoted by I.

If we now apply this definition to the whole rod considered, we meet a difficulty. The condition of the material, in regard to magnetization, is evidently not the same everywhere in the rod, as is shown by the fact that the poles are not exactly on the ends of the rod. We must evidently apply the definition to a small part of the rod at a time to get the intensity of magnetization of the small part. For this purpose we consider a small cylindrical portion of the bar, which is uniformly magnetized parallel to its length. For such a cylinder magnetic poles would appear only on the ends, none on the sides.

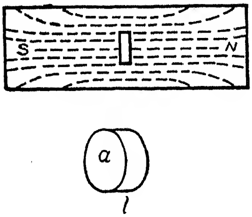


FIG. 343.

Suppose that the length of the cylinder is l, its pole strength m, its magnetic moment M, its area of pole a, and its volume v. Then by definition of I

$$I = \frac{M}{v} = \frac{ml}{al} = \frac{m}{a}$$

Thus intensity of magnetization is also *pole strength per unit area.* This second definition of I, like the first, does not mean much when applied to the whole rod; but, when applied to the small specimen we have considered it has a definite meaning.

440. Magnetic Induction.—A second term is found necessary in describing the state of magnetized materials. To see how it comes in, consider the case of a rod of soft iron magnetized by a magnetic field inside a solenoid. If the iron is not in the solenoid, there is inside the solenoid a certain magnetic field strength H represented by H lines of force per unit area. When the iron is introduced the total number of lines of force is found to be greatly increased. Evidently the presence of the iron produces additional magnetic lines running through the iron and emerging into the air. The lines of force from the solenoid and the iron are collectively spoken of as lines of *magnetic induction* or *magnetic flux.*

In an earlier section (§408) a magnetic field of H oersteds was represented by H lines of force per cm.^{2} perpendicular to the field. The number of lines of magnetic flux per cm.^{2} is called the *magnetic flux density* and is measured in *gausses.* The usual symbol for flux

density is B. If there is a uniform flux density of B gausses over an area of A cm.^{2} then the total flux ϕ through the area A is BA and is measured in *maxwells*.

$$\phi = BA \qquad\qquad B = \frac{\phi}{A}$$

Maxwells = gausses × sq. cms. Gausses = maxwells per sq. cm.

To find the magnetic flux within the iron we must go back to the definition of intensity of magnetization I. Consider a small pill-box shaped cylinder cut out of the iron with its ends perpendicular to the direction of alinement of the elementary magnets, that is, perpendicular to the direction of magnetization I. The strengths of the poles that appear on the ends of this cylindrical cavity are the same as those appearing on the ends of the cylinder which is cut out. If I is the intensity of magnetization in the region considered, then the pole strength per cm.^{2} appearing on the ends of the cylindrical cavity is I. The total number of lines of force or magnetic flux within the cavity is the sum of those due to the magnetizing field H and those due to the poles on the ends of the cavity. Now to find the number of these lines due to the pole faces is a problem exactly similar to the one we had in finding the number of electric lines of force per cm.^{2} between two parallel charged plates (§391), except that instead of Q/A, charge per unit area, we now have m/a, pole strength per unit area. The number of lines of magnetization per unit area is therefore $4\pi m/a$ or $4\pi I$.

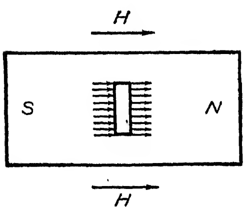


FIG. 344.

The total number of lines of force per unit area is called the magnetic *flux density* or magnetic induction B, so that

$$B = H + 4\pi I$$

The lines of magnetic flux are continuous, passing out of the iron at the N pole of the bar into the air and returning at the S pole. Now H, the magnetic field within the iron, which gives rise to B, is usually less than the external field because of the N and S poles induced on the ends of the bar of iron. These magnetic poles produce a field in a direction opposite to that of the magnetizing field and hence reduce it. This demagnetizing effect, as it is called,

is almost impossible to calculate exactly, but its effect is negligible in a bar that is long compared to its diameter. Therefore for a very long thin bar magnetized in a field parallel to its length, the magnetic field H within the bar may be taken as equal to the external field.

441. Permeability and Susceptibility.—The terms intensity of magnetization and magnetic induction are descriptive of the *state* of a magnetized substance, just as density refers to a particular state of a compressible substance. The two terms we are now to introduce relate to the *nature* of the substance, how far it is magnetizable, just as the compressibility of a substance means how readily its density can be changed.

The ratio of magnetic flux density B in a substance to the strength H of the magnetic field that produces B is called the *permeability* of the substance and is denoted by μ. Hence

$$\mu = \frac{B}{H} \qquad \text{or} \qquad B = \mu H$$

The ratio of intensity of magnetization I to the field intensity H is called the *susceptibility* of a substance and is denoted by k. Thus

$$k = \frac{I}{H} \qquad \text{or} \qquad I = kH$$

The relation between μ and k can be found by writing the equation connecting B, H and I (§440) in the form

$$\frac{B}{H} = 1 + 4\pi\frac{I}{H}$$

Hence

$$\mu = 1 + 4\pi k$$

Thus, if either μ or k can be measured, the other can be calculated. As regards the most important magnetic substance, iron, it is usually I that is measured. Different methods of measuring it have been used. For the usual one see §501.

Permeability μ was introduced in §405 with reference to the force between the two magnetic poles, $F = m_1m_2/\mu d^2$, the force being reduced when the poles are immersed in a medium of permeability μ as compared to a vacuum. While it is not possible to show here that the μ introduced in the two connections is the same, it is possible to see qualitatively how the magnetic medium reduces

the force between the isolated poles. The magnetic poles have the effect of aligning the elementary magnets of magnetic material in such a manner as to reduce this force. A N magnetic pole draws the S poles of the elementary magnets towards it, thus effectively reducing its own strength (compare p. 353).

442. Effects of High Permeability. Magnetic Shielding.—When a magnetic substance is placed in a magnetic field, it changes the distribution of lines of force. This is shown by the arrangement of iron filings about an iron disk placed in a uniform field (Fig. 345). The lines tend to go through the iron because of its higher permeability. Fig. 346, taken from a paper by Kelvin, shows the field around and through a sphere of high permeability.

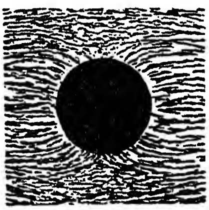


FIG. 345.

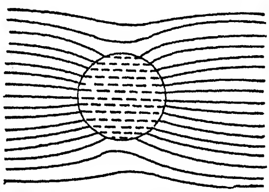


FIG. 346.

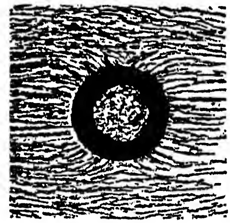


FIG. 347.

Fig. 347 shows the lines about a thick iron ring in a uniform magnetic field. It is to be noticed that the lines are diverted toward the ring on account of its higher permeability; and also that the filings on the inside of the ring show little, if any, directive action. This "screening" effect of soft iron is made use of in protecting sensitive galvanometers from magnetic disturbances. The same principle is used in shielding the interior parts of the coils of the ring armature of a direct current generator (Fig. 423).

443. Magnetic Circuit.—Maxwell's term *magnetic flux* or flow of induction has led to the very useful idea of a *magnetic circuit*, analogous to an electric circuit in which an electromotive force produces a flow of electricity and is opposed by an electric resistance. The law of this circuit is analogous to Ohm's law (§420) for an electric circuit. As the simplest case consider the toroidal ring of §431. $H = 4\pi Ni/l$, where N is the whole number of turns of the solenoid. If S is the cross-section of the core, the magnetic flux $= BS = H\mu S = 4\pi Ni \div l/\mu S$. In this, $4\pi Ni$ is called the *magnetomotive force* and $l/\mu S$ the magnetic resistance or reluctance of the circuit. Hence

$$\text{magnetic flux} = \frac{\text{magnetomotive force}}{\text{reluctance}}$$

The magnetomotive force $= 4\pi Ni = 4\pi$ multiplied by Ni which is called the "ampere turns." The reluctance is proportional to $1/\mu$, which is called the

specific reluctance of the material; to l, the length of the magnetic circuit; and inversely to S, the cross-section of the core. This relation can be extended to complex magnetic circuits in which there are air gaps, such as an electromagnet or the field magnet of a dynamo. The total reluctance is then the sum of parts due to the separate materials (iron and air), but the results are only rough approximations.

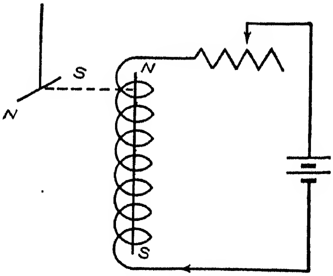


Fig. 348.

444. Residual Effects. Hysteresis. Starting with an apparatus as shown in Fig. 348, suppose the iron, wholly unmagnetized, to be in the solenoid before the current is applied. When the current is started the iron becomes magnetized, as shown by a deflection of the magnetic needle. If the current be now broken the iron will not return fully to its initial unmagnetized condition; it will remain partly magnetized. This is called a *residual magnetic effect.* A similar result is obtained if the magnetizing field, already in existence, is increased so as to produce an increase of magnetization and is then decreased to its first value. The increase of magnetization is not fully removed; some of it remains. These residual effects vary greatly with the nature and state of the iron. Generally speaking, they are greater the harder the iron. They are small for soft iron and large for steel. This property of retaining magnetism is called *retentivity.*

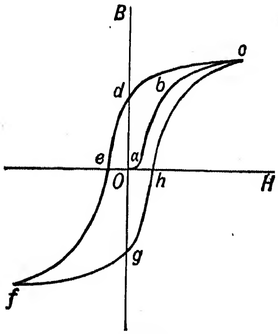


Fig. 349.

When alternating currents are applied to an electromagnet of any form, the iron goes through cycles of magnetization. Because of the retentivity of the iron, magnetization at each stage lags behind the magnetizing field applied. This effect is called *hysteresis* (lagging). For any specimen of iron it is studied by constructing a *hysteresis loop* (Fig. 349) with applied magnetic field strength H as abscissa and magnetic induction B as ordinate. The magnetizing field H is first increased from zero by steps and the increase of B at each step is found. If the iron is initially unmagnetized these observations give the curve *abc*. H is next decreased, step by step, to zero. The result is the curve *cd*, where *ad* represents the residual

induction or remanence. H is then reversed, and is then increased, step by step, in the reverse direction. At some value of H, B is reduced to zero as shown by the point e. The magnitude of H reversed that is required to reduce B to zero is called the *coercive force* of the specimen. H is next increased in the reverse direction up to a negative maximum equal in magnitude to the positive maximum. This completes half of the loop. The other half *fghc* is found, step by step, by first decreasing H in magnitude and then reversing it and continuing up to the positive maximum represented by the point c.

Heat is always produced in hysteresis owing to internal friction in changing the orientation of atoms or atomic aggregates, and the area of the hysteresis loop is a measure of the heat so produced. The harder the iron, the greater is the area of the loop and the larger the amount of heat generated. Iron can be treated so that, for a maximum induction of 100 gausses, the heat produced per cm.^{3} per cycle is only 0.1 erg. A magnetic alloy, perminvar, consisting of iron, nickel and cobalt, has a heat loss of only 0.00003 erg per cm.^{3} per cycle at the same maximum induction.

Studies of permeability and hysteresis in different kinds of iron were first made by a magnetometer method, similar in principle to that sketched in Fig. 348. A more convenient and more accurate method, the one now mostly employed, depends on the use of induced currents (§501).

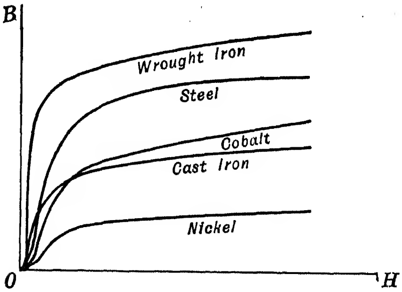


FIG. 350.

445. Different Magnetic Substances.—While iron is the most magnetic substance, having the highest permeability, cobalt and nickel are also very magnetic. These three elements and certain alloys of one or more of them, containing also such substances as chromium, manganese, tungsten, etc., are classed together as *ferromagnetic* substances. The magnetic properties of any one of them vary widely with its condition, especially with its temperature. At about 770°C. iron suddenly loses its exceptional magnetic properties and, while still solid, becomes only feebly magnetic. This is also the temperature of recalescence, as it is called, the temperature, that is to say, at which a piece of hot iron that is cooling suddenly

glows brightly, thus showing that some important rearrangement of atoms or atomic groups has occurred, accompanied by the liberation of energy.

Fig. 350 shows in a very general way the magnetic properties of different kinds of iron and of cobalt and nickel. The table below

H	I	$k = I/H$	B	$\mu = B/H$
0	0	...	0	
0.32	3	9	40	120
0.84	13	15	170	200
1.37	33	24	420	310
2.14	93	43	1,170	550
2.67	295	110	3,710	1,390
3.24	581	179	7,300	2,250
3.89	793	204	9,970	2,560
4.50	926	206	11,640	2,590
5.17	1,009	195	12,680	2,450
6.20	1,086	175	13,640	2,200
7.94	1,155	145	14,510	1,830
9.79	1,192	122	14,980	1,530
11.57	1,212	105	15,230	1,320
15.06	1,238	82	15,570	1,030
19.76	1,255	64	15,780	800
21.70	1,262	58	15,870	730

gives an example of the experimental data from which such B-H curves, as they are called, are drawn. The figures show how rapidly the permeability drops off after the appearance of the bend of the curve that indicates the approach of saturation. It is to be remembered that μ always means B/H, not changes of B divided by changes of H. The curves and table show that the permeability μ of iron is not constant but depends on the magnetic field and the previous history of the iron.

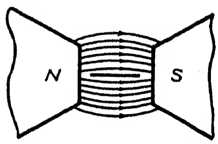


FIG. 351.

446. Paramagnetism.—The ferromagnetic substances, iron, nickel, and cobalt, are attracted by a magnet and they have permeabilities greater than unity (usually very much greater). Many other substances have similar properties but are only very feebly attracted by a magnet and have permeabilities only slightly greater than unity. They are described as *paramagnetic*. This class includes about twenty metallic elements such as aluminum, chro-

mium, manganese, sodium, and platinum. Oxygen is paramagnetic as a gas, liquid, or solid.

The properties of paramagnetic substances are of theoretical interest but of no practical importance. So far as tested, they do not exhibit saturation of magnetization or hysteresis, possibly because sufficiently strong fields have not been applied. Iron above the critical temperature, 770°C., is paramagnetic and the same is true of the other ferromagnetic substances above their critical temperatures.

447. Diamagnetism.—A diamagnetic substance is repelled by a magnet and its permeability is less than unity. A rod of a ferromagnetic substance in a magnetic field tends to turn parallel to the field and to move into the strongest part of the field (Fig. 351), but a rod of a diamagnetic substance in a magnetic field tends to turn at right angles to the field and to move into the weakest part of the field (Fig. 352). Bismuth, the most strongly diamagnetic substance, has a permeability of 0.9998. Antimony, phosphorus, quartz, glass, copper, silver, water, in fact the great majority of substances, including nearly all gases, are diamagnetic, but the property is so weak that very special means have to be devised for detecting it at all.

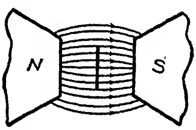


FIG. 352.

Diamagnetism appears to be due to a modification of the motions of electrons within atoms, produced by a magnetic field, but the subject is too complex for discussion here.

448. Three Stages in the Magnetizing Process.—When a series of values of H and I for a specimen of soft iron has been obtained and plotted as a curve (Fig. 353), it is seen that it consists of three fairly distinct parts. First (O to A) there is a slow increase of I as H is increased, which initially is linear; then (A to D) a very rapid increase; and finally (after D) another slow increase, that indicates an approach to a maximum.

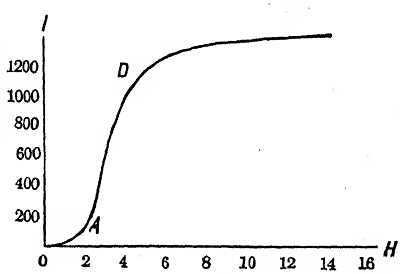


FIG. 353.

These peculiarities can be explained by regarding the soft iron, before it is magnetized, as consisting of a great number of very

small elementary magnets, or "domains," with their axes turned in random directions, so that no external effects are produced. Each domain is believed to consist of a group of a great many molecules, each magnetic, and with their magnetic axes in a common direction. The first effect of the magnetic field is to make the axes of the molecules in each domain turn slightly toward the direction of the field. This accounts for the part OA of the curve. With a stronger field all the molecules in a domain wheel suddenly from their original position of stability to a new position, so that their axes will be more nearly in the direction of the field. This takes place in different domains in rapid succession, which corresponds to AD. When all of them have wheeled around, the molecules again turn slightly

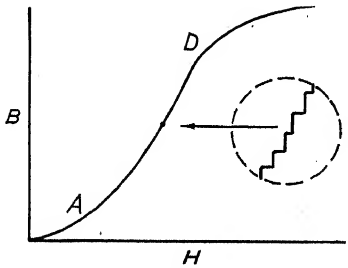


FIG. 354a.

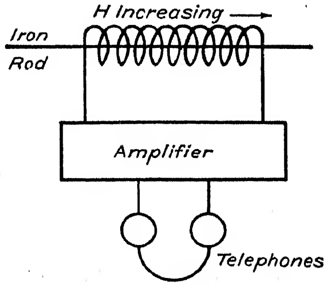


FIG. 354b.

toward the direction of the field, but this produces only a small increase in B.

There is some evidence for this account. When a weak field is applied slowly, the initial part of the curve is linear and reversible, as would be expected if the molecular axes were able to turn back when released. With higher magnetization (A to D) the effect is not reversible; when the current is broken there is some permanent magnetization. Now in this stage, if the field is applied slowly, each sudden wheeling of a domain will produce an electrical impulse in a coil connected to the input of an amplifier (Fig. 354*b*), as evidenced by an audible click of the output loudspeaker; these clicks occur in rapid succession (Barkhausen effect). There is evidence that each domain, while invisibly small, consists of something like 10^{15} atoms. But why should the atoms be magnetic? Ampere gave an answer and modern research has upheld and amplified it.

449. Atoms as Magnets.—As an explanation of the magnetic properties of bodies Ampere suggested that the ultimate atoms are themselves magnetic because of the existence of electric currents in the atoms, the magnetic fields of the currents giving the atoms the properties of magnets (§425). While this theory of amperian currents has always been thought interesting, until recently there was no other evidence for their existence. Now, however, there are strong reasons, chiefly of an optical nature, for believing that atoms consist of electrons revolving about nuclei (§366), or, at least, that the state of affairs in an atom is so well represented by such a picture that an immense number of mathematical deductions from it are found to be verified by very exact measurements. An electron moving in an orbit and completing a revolution many times in a second is equivalent to a current and provides the current assumed in Ampere's theory. The following is a brief sketch of how the idea was developed by Niels Bohr, Danish physicist, in 1913.

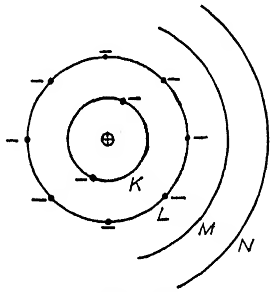


FIG. 355*a*.

An atom of any substance consists of a *nucleus* that has a net positive charge and a number of *electrons* (negative charges) in orbital motions around the nucleus at distances that are very large compared with the dimensions of nucleus or electron (Fig. 355*a*). What the nucleus consists of we shall consider later (§556). At present we are interested in the motions of the electrons. Their number, called the *atomic number* of the element, is such that their aggregate charge just balances the net positive charge of the nucleus, leaving the normal atom (not an ion) electrically neutral. The atomic number of hydrogen is 1, that of uranium 92, while the atomic numbers of all other elements are between these extremes, that of oxygen, for example, being 8, of iron 26, of lead 82, and so on.

The electrons surrounding a nucleus are not distributed in any random way but are grouped in "shells." The helium atom has one shell, consisting of only two electrons and called the *K*-shell (Fig. 355*a*). The next heavier atom, that of lithium, also has a *K*-shell, but in addition it has a third orbital electron, at a greater distance from the nucleus, that begins a second shell, called the *L*-shell. In

succeeding atoms the number of electrons in the L-shell grows, one by one, until in the neon atom there is, in addition to a K-shell, a completed L-shell of 8 electrons. Next comes sodium with a complete K-shell and a complete L-shell and an additional electron that begins an M-shell. When fully completed in any atom this M-shell contains 18 electrons. But before it is completed in successive atoms, something different may take place, and this feature is what interests us in connection with magnetism. We may illustrate it by the case of iron.

The atomic number of iron is 26. It has a K-shell of two electrons, an L-shell of 8 electrons, and 14 electrons in an incomplete M-shell. What has become of its other two electrons? They have gone to start an N-shell, before completing the M-shell. Now in elements that show the strongest magnetic effects, such as iron, manganese and nickel, *electrons go into an outer shell before the preceding shell is complete.* There are two ways in which the electrons may then give to the atom the properties of a magnet:

(1) Any complete shell of electrons is magnetically neutral; for, while each revolving electron has a magnetic field, the orbits are so distributed in planes around the nucleus that they cancel out as regards magnetic effects. But *an incomplete shell has a magnetic field;* iron, for example, has an incomplete M-shell and an incomplete N-shell;

(2) In addition to revolving around the nucleus, electrons also *spin on axes of their own,* just as the earth rotates on its axis, while revolving about the sun. In any atom the axes of spin of all the electrons are parallel, but, within a single atom, the spins may be in either of the two opposite directions around the axis. Now *a spinning electron has a magnetic field due to the spin.* For iron the positive and negative spins are as indicated in Fig. 355b by + and − signs, so that there are four uncompensated or excess electron spins contributing to the magnetic moment of the atom. There is also a spin of the nucleus, but this is of minor importance.

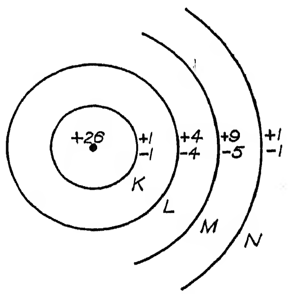


FIG. 355b.

The magnetic moment of the atom, when it is free from external influence, is the resultant of these two parts. When it is in a magnetic field, a spinning electron is subject to a couple that causes it

to precess like a top or gyroscope (§129). In fact, it has been found possible, by suspending a light body of magnetic material inside a solenoid and turning the current on and off, to get a reaction or kick that can only be accounted for by precession of the spinning and revolving electrons. Also the external magnetic field causes a shift of the planes of the orbits and some modification of the dimensions of the orbits.

While the effects of these different actions are complex, they can be worked out for different types of atoms and an explanation of magnetic and diamagnetic properties can be found, but the details cannot be given here. Other phenomena that can be explained by Bohr's model of the atom will be referred to later (§§540, 729).

RESISTANCES OF SOLIDS

450. Resistance and Resistivity.—The definition of electrical resistance as E/i (§420) provides a direct method for measuring it, namely by finding E by a voltmeter and i by an ammeter, and calculating E/i. An arrangement for making the measurement is shown by Fig. 356. A battery B produces a current i in a resistance R and the current is measured by an ammeter A. The electromotive force E or potential drop between the ends of R is measured by a voltmeter V. If E is in volts and i in amperes, this gives R in ohms. One assumption in the method is that no appreciable part of the current read by A goes through V; but the resistance of a good voltmeter is so high that, for ordinary work, the assumption leads to no appreciable error. The method also depends on the accuracy of graduation of ammeter and voltmeter and this may, in practice, be a considerable source of error. For the accurate work which is sometimes necessary a method of comparison of resistances is preferable and one will be considered later (§453).

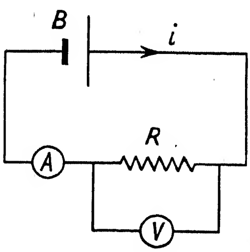


FIG. 356.

Experiments of a wide range have shown that there are two simple relations between the resistance of a conductor in the form of a wire and its dimensions. The resistance of a uniform wire is (*a*) proportional to its length l and (*b*) inversely proportional to its cross-section A. These relations are what would be expected from the

theory of the nature of conduction and resistance stated earlier (§418).

We may therefore write for the resistance of a length l of a uniform wire, or cylindrical conductor, of cross-sectional area A

$$R = \rho\frac{l}{A}$$

We now note that the constant of proportionality ρ, which is different for different materials, is what R becomes if $l/A = 1$. For any substance it is the resistance of a wire or rod such that its length divided by its cross-section is 1. It is called the specific resistance or *resistivity* of the material and may be defined as *resistance per unit length per unit reciprocal of area of cross-section.* We may also say that the specific resistance of a substance is the resistance of a cube of the substance of 1 cm. edge (Fig. 357), with the understanding that the current is perpendicular to two opposite faces and is distributed uniformly over them; but it is *not* resistance per unit volume.

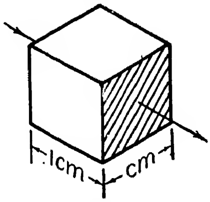


Fig. 357.

As the reciprocal of the resistance of a conductor, *i.e.* $1/R$, occurs frequently, it is desirable to have a name for it. It is called the *conductance* of the conductor. Similarly the reciprocal of resistivity is called *conductivity.*

Specific Resistance

Material	Specific resistance in ohm × cm. at at 20°C.	Temperature coefficient per °C.
Aluminum	2.8×10^{-6}	0.0039
Copper	1.7×10^{-6}	0.004
Iron	10.0×10^{-6}	0.005
Manganin	44×10^{-6}	0.00001
Mercury	95.8×10^{-6}	0.00089
Platinum	10.0×10^{-6}	0.0036
Nichrome	100×10^{-6}	0.0004
Silver	1.6×10^{-6}	0.0038
Carbon	3500×10^{-6}	−0.0005

451. Physical Conditions Affecting Resistances.—The electrical resistance of a body is constant only so long as its physical condition

is unchanged. Any change in physical condition gives the body a new resistance different from what it had before. Impurities have often a very marked effect. Changes in hardness and mechanical strains affect it. The resistances of some substances, such as selenium, are affected by exposure to light; others, such as bismuth, show a change of resistance in a magnetic field. One of the most marked effects is that of change of temperature, and we shall consider it separately. These influences modify in some way the number or condition of the free or conduction electrons on which conduction and resistance depend.

As any change in resistance of a wire can be used to affect a current and so change the reading of some electrical instrument, many of the influences we have mentioned have found important practical applications. The mention of one is appropriate here; a spiral of bismuth, mounted so that its resistance affects the reading of a galvanometer, can be used as a means of measuring the strengths of magnetic fields.

452. Resistance and Temperature.—The electrical resistance of pure metals increases as their temperature is raised. The expression for the relation is similar to that for the expansion of a bar when heated. Thus if r_0 is the resistance of a wire of a pure metal at 0°C. and r is its resistance at t°C.,

$$r = r_0(1 + at)$$

where a is the *temperature coefficient of resistance* of the metal over the temperature range from 0°C. to t°C. Actually a is not exactly a constant; it is really an average value, but its variation over a moderate range of temperature may be neglected. It is interesting to note (see Table) that for pure metals a is of the same order of magnitude as the coefficient of expansion of a gas ($\frac{1}{273} = 0.00366$), which seems to indicate that free electrons in a metal have, to some extent, the properties of a gas, though difficulties have been found in the attempt to carry this analogy farther. Carbon has a negative temperature coefficient, that is, its resistance decreases with rise of temperature.

Standards of resistance should be made of substances with very small temperature coefficients. An alloy called manganin, consisting of 84 per cent copper, 12 per cent manganese, and 4 per cent nickel, is best for the purpose, as it has an extremely small temperature coefficient.

Since the resistance of a coil of wire is a function of its temperature, the relation can be used for the measurement of temperature. Platinum has been found the most satisfactory metal for resistance thermometry, as it is called. The advantages of a platinum thermometer are great sensitiveness and the wide range of temperatures, extending from the lowest up to about 1200°C., for which it can be used.

453. Wheatstone Bridge for Comparing Resistances.—In the arrangement of Fig. 358 X is an unknown resistance and R_1, R_2, R_3 are known resistances that can be varied in magnitude. A galvanometer is used for detecting whether there is a potential difference between B and C when a battery is connected between A and D. R_1, R_2, R_3 are varied until there is no current through the galvanometer when the key K is pressed. The Wheatstone bridge is then said to be "balanced."

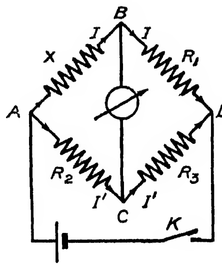


Fig. 358.

To find the relation between the resistances when the bridge is balanced, we note that, since there is no current through the galvanometer, the current in AB is equal to that in BD, say I; and the current in AC is equal to the current in CD, say I'. Also B and C are at the same potential; hence the potential drop from A to B, which is IX, is equal to that from A to C, which is $I'R_2$. Hence

$$IX = I'R_2 \tag{1}$$

Similarly the potential drop from B to D is equal to that from C to D. Hence

$$IR_1 = I'R_3 \tag{2}$$

Hence by division

$$\frac{X}{R_1} = \frac{R_2}{R_3} \tag{3}$$

From this equation X can be found if R_1, R_2, R_3 are known.

This equation for a balance can also be written in the form:

$$XR_3 = R_1R_2 \tag{4}$$

that is, for a balance, the products of the resistances of opposite arms must be equal. This shows that, when a balance has been obtained, there will still be a balance if the positions of galvanometer

and battery are interchanged in Fig. 358; for the necessary condition, equation (4), is the same in both cases.

454. Forms of Wheatstone's Bridge.—In the oldest and simplest apparatus for measuring resistance by the Wheatstone's bridge method, called the "slide-wire" bridge (Fig 359), two of the arms

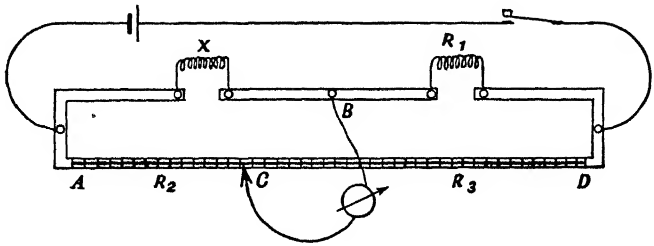


FIG. 359.

R_2 and R_3 are the parts of a uniform wire, usually a meter long and provided with a scale divided to mms. The ratio R_2/R_3 is then the ratio of the lengths of the two parts of the wire, and only a single "known" resistance, R_1, is needed. (The resistances of the heavy rectangular strips of copper are regarded as negligible.) There is, however, always some doubt as to the uniformity of such a wire, and as to whether its two ends are attached in just the same

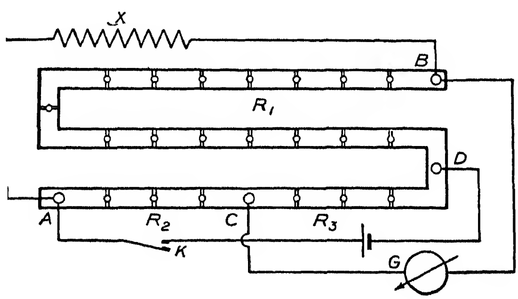


FIG. 360.

way. A more reliable result is obtained by taking two readings, X and R_1 being interchanged, and averaging.

For more accurate work a bridge box is used (Fig. 360). It contains coils of 1, 10, 100 ohms in each of the two "ratio arms," R_2 and R_3, and a series of resistances running up to 5,000 ohms for the "known arm," R_1. With the combinations possible, this gives a range from $\frac{1}{100}$ of 1 ohm to 100 times 5,000 ohms.

455. Measurement of Very High Resistances.—A method that can be used is to find at what rate a charge in a condenser leaks through the high resistance, when it is connected between the plates of the condenser (of much higher internal resistance). If the capacity of the condenser is C and its initial charge Qo, after time t its charge falls to Q, where

$$Q = Q_o e^{-\frac{t}{XC}}$$

e being the base of Naperian logarithms and X the unknown high resistance. If the charges Qo and Q are measured by discharge through a ballistic galvanometer (§437) and t observed, X can be calculated. A mica condenser is usually employed, for mica as a dielectric does not retain a residual charge.

The equation can be derived by noting that, when the charge is Q, the difference of potential of the plates of the condenser is Q/C and the potential drop in the resistance is Xi, which is equal to $-XdQ/dt$. By equating these and integrating we get the equation stated.

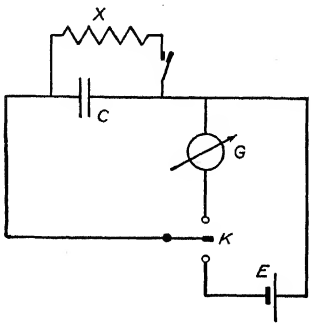


FIG. 361.

456. Resistance Standards and Resistance Boxes.—An exact measurement of resistance is of such great importance in many practical affairs that reliable standards to be used in the measurement are necessary, just as reliable weights are necessary for weighing. Fig. 362 shows a form in which a wire that has as exactly as possible a resistance of 0.1 ohm is mounted. Standards of 0.01 ohm and 0.001 ohm are also used. As it is especially important that such standards vary as little as possible with temperature, they are made of some alloy, such as manganin, that has a very low temperature coefficient. In this Bureau of Standards type of fixed standards, as they are called, the coil is sealed into an oil-filled vessel and the temperature of the coil can be read by a thermometer.

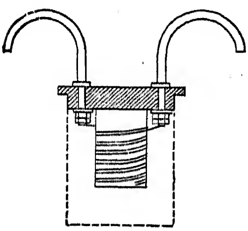


FIG. 362.

A resistance box (Fig. 363) corresponds to a box of weights. It has a graduated series of resistance coils, attached to heavy brass blocks that are mounted on the cover of the box, which is of a good insulating material, such as ebonite. All the coils are connected in series, but a brass plug inserted between the two blocks to which a coil is attached offers so little resistance that the coil is

"cut out." Thus any series combination of the resistances up to their sum can be used. This is called a plug box. In a different form of box, called a dial box, all the resistances in a decade are equal and a rotating arm or switch can throw any sequence of them into the circuit. The coils are of manganin or some alloy with a very low temperature coefficient, and each is wound back on itself to eliminate effects of self-induction (§492).

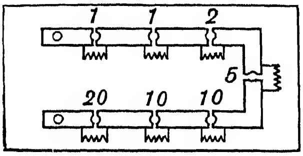


FIG. 363

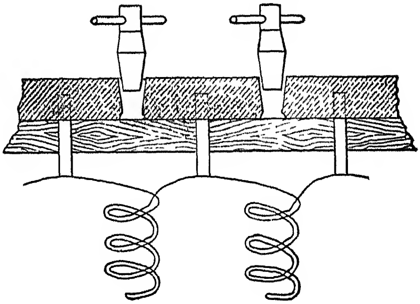


FIG. 364

Rheostats or variable resistances for controlling currents are made in a variety of forms. In one, bare wire is wound on an insulating cylinder, and a slider can be pushed along the coils so as to include as many of them as necessary in the circuit. In the carbon rheostat the current passes through a series of carbon disks and the resistance depends on pressure applied to force them into closer contact.

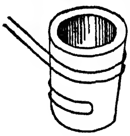


FIG. 365.

457. Resistances of Combinations of Conductors.—Conductors can, of course, be combined in a great variety of ways. There are, however, two fundamental combinations into which others can be analyzed.

(*a*) *Series Arrangement.*—Conductors are in series when they are joined so that the same current passes through all (Fig. 366). It will be sufficient to consider the case of three conductors of resistances, R_1, R_2, R_3. When a current I passes through them, the

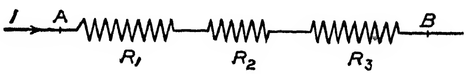


FIG. 366.

potential drops in them are IR_1, IR_2, IR_3. If R is the single resistance that is equivalent to the series, it can be substituted for the series, leaving both drop of potential and current unchanged. Hence

$$IR = IR_1 + IR_2 + IR_3$$

and therefore

$$R = R_1 + R_2 + R_3$$

When conductors are in series the total resistance is the sum of the resistances.

(*b*) *Parallel Arrangement.*—Conductors are in parallel when they are joined so that the potential difference across them is the same and the whole current I is divided into parts I_1, I_2, I_3 in the conductors. Then

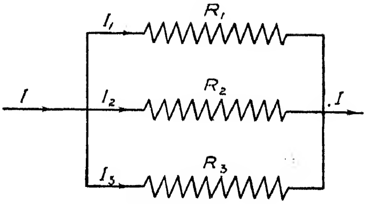


FIG. 367.

$$I = I_1 + I_2 + I_3$$

If the resistances are R_1, R_2, R_3 and if R is the equivalent single resistance, each current is equal to the potential drop, divided by the corresponding resistance. Substituting we get

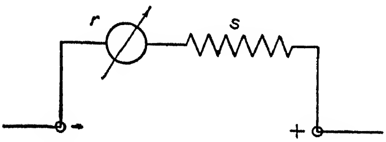


FIG. 368.

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

When conductors are in parallel, the reciprocal of the total resistance is equal to the sum of the reciprocals of the resistances.

Briefly we may say: in the series arrangement resistances are additive, in the parallel arrangement conductances are additive, and other arrangements can usually be resolved into these simple ones.

458. Increased Range of Voltmeter and Ammeter.—A voltmeter of resistance r and graduated for a certain range of voltage can be used for a wider range by using a series resistance s with it. (Fig. 368). For example, if the range is to be increased 10-fold, s must be such that $r + s = 10r$, or $s = 9r$. It is readily shown that if the range is to be increased n-fold, $s = (n - 1)r$.

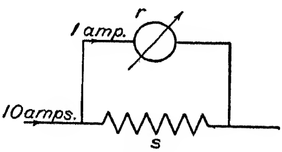


FIG. 369.

The range of an ammeter of resistance r can be increased by a shunt of resistance s (Fig. 369). If the range is to be increased 10-fold, $\frac{1}{10}$ of the whole current is to go through the ammeter and therefore $\frac{9}{10}$ through the shunt. Hence $s = r/9$. If the range is to be increased n-fold, $s = r/(n - 1)$.

459. Cells in Series and in Parallel.—Cells are in series when the positive terminal of each is connected to the negative of the next (Fig. 370). If n cells, each with an e.m.f. of e volts and an internal

resistance of r ohms, are connected in series, their total e.m.f. is ne volts and their total resistance is nr ohms. If they are in a circuit with R ohms external resistance and give a current of I amperes, by Ohm's law

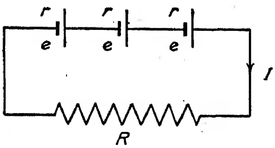


FIG. 370.

$$I = \frac{ne}{R + nr}$$

Cells are in parallel when all the positive terminals are connected together and all the negative connected together (Fig. 371). If the e.m.f. of each cell is e and its resistance r and there are n in parallel, the e.m.f. is e and the total internal resistance is r/n. The current I they give when connected to an external resistance R is:

$$I = \frac{e}{R + \frac{r}{n}}$$

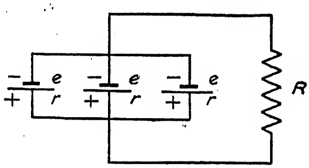


FIG. 371.

The current through each cell is I/n.

In general the purpose of the series arrangement is to get a high e.m.f. and that of the parallel arrangement to get a large current without danger of injuring a cell. When cells are connected in parallel it is necessary that they have the same e.m.f. and internal resistance.

460. Kirchhoff's Laws.—For calculating steady currents in a complex network it is very important to have some systematic standard procedure, something like the equations that can be written down for finding the forces in a framework. Such a general method of attack is provided by two laws formulated by Kirchhoff. While they do not add any new fact or principle they enable us to write down at once a sufficient number of equations for finding the currents in a network, so that the rest is a matter of algebra.

1. *The algebraic sum of the currents that meet at a point is zero.* *Sum* $I = 0$.

2. *In any closed path the algebraic sum of the IR terms is equal to the algebraic sum of the applied e.m.f.'s.* *Sum* IR = *Sum* E.

The first law is equivalent to saying that, when the currents in a network are steady, as much electricity flows out at a junction as flows in. This means that the number of conduction electrons does not change and they do not accumulate at any point.

In applying the second law to any closed path in the network one direction around the path (it does not matter which) is to be taken as positive. A potential drop in that direction in a conductor, or an applied e.m.f. (a cell or any generator) that tends to send a current in that direction, is to be given a positive sign, any in the opposite direction being called negative. Since there

a rise of potential in a cell (or any generator) in the direction in which it ends to send a current, the law simply means that in a closed path the sum of the rops is equal to the sum of the rises of potential.

For example, consider the arrangement in Fig. 372. The actual directions f the currents I_1, I_2, I_3 are not known initially; so they are tentatively taken as ndicated by the arrows. Then from the two laws:

At A or C $\quad I_1 - I_2 - I_3 = 0$

Around $ABCDA$ $\quad 4I_2 - 6I_3 = -2 - 4 = -6$

Around $ABCEA$ $\quad 4I_2 + 10I_1 = -2 + 6 = +4$

These equations can be solved for I_1, I_2, I_3. They give for I_2, for example, $-\frac{9}{31}$ ampere, which shows that this urrent is in the opposite direction to he arrow.

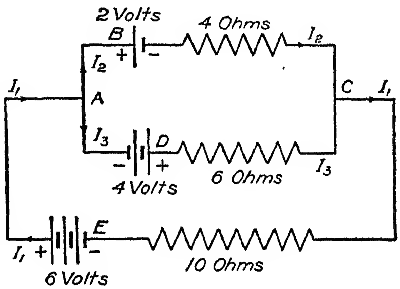


Fig. 372.

461. The Potentiometer.—The ratio of two potential drops or the ratio of the electromotive forces of two cells can be found by means of an arrangement called a *potentiometer*. We shall suppose that the purpose is to compare two cells of e.m.f. E and E' (Fig. 373). A battery B sends a constant current through a uniform wire AD in the direction from A to D, so that along AD there is a uniform potential drop. The two cells are connected to a double-throw switch so that either can be connected for a moment through a galvanometer G to A and to a point C on AD. If the positive pole of the cell is connected to A, the cell and the difference of potential at A and C tend to send currents in opposite directions through G. If the whole potential drop in AD is greater than E or E', a point C for E and a point C' for E' can be found so that there is no current in the galvanometer. In each case the potential drop is equal to the e.m.f. of the cell. Hence if the lengths of AC and AC' are l and l'

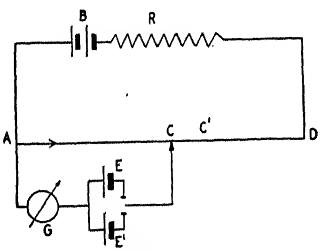


Fig. 373.

$$\frac{E}{E'} = \frac{l}{l'}$$

There is one adjustment of the potentiometer that is especially convenient. If one of the cells is a Weston standard cell, the e.m.f. of which is 1.0183 volts at 20°C., the resistance R can be adjusted so that the length AC for balancing

the standard cell is 101.83 cm. The drop per cm will then be 0.01 volt and E' will be given by the length AC' in meters. For work of the highest accuracy the wire is replaced by a series of resistances in a closed box, and this, with the necessary switches, is called a potentiometer.

Since there is no current in the galvanometer and cell when a balance is obtained, the potential drop in the wire is a measure of the maximum potential difference of the poles of the cell, that is, its e.m.f. If a voltmeter were connected directly to a cell to get its e.m.f. there would be a potential drop, IR, within the cell due to its internal resistance R; and this would have to be added to the potential drop in the voltmeter to get the true e.m.f. of the cell.

462. Use of a Potentiometer to Measure a Current.—While a potentiometer is primarily used for the measurement of potential differences, it can also be used to measure a current by finding the potential drop in a known resistance in the circuit of the current. It can, for example, be used to test the graduations of an ammeter. For this purpose the cell E' in Fig. 373 is replaced by the circuit shown in Fig. 374, where the known resistance r is in series with the ammeter, a battery and a rheostat R. The potential drop Ir in r is found by the potentiometer, and this gives I. By varying the current by means of R the whole scale of the ammeter can be calibrated.

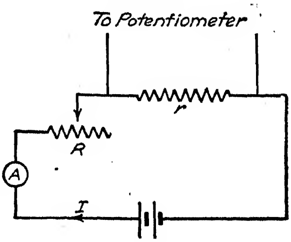


FIG. 374.

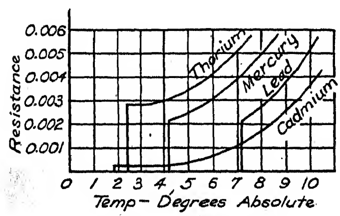


FIG. 375.

463. Conduction near Absolute Zero.—When the temperature of a metal is lowered very greatly, even to within 10° of absolute zero (−273°C.), the resistance steadily decreases, though not as rapidly as would be represented by the linear relation $r = r_0(1 + at)$. But at a few degrees from absolute zero, as discovered by Kammerlingh Onnes, the resistance of some metals, not of all, suddenly falls practically to zero (Fig. 375). For lead this occurs at 7.2° above absolute zero, for mercury at 4.2°, for tin at 3.8°, for thallium at 2.3°. This is called *superconductivity*.

A current can be started in a superconducting ring by drawing away a magnet placed at the center of the ring (§483). The ring shows the presence of a current by the deflection of a compass needle, and the current goes on flowing,

with a decrease of perhaps 1 per cent per hour, so that it can be detected after several days; in fact in one case it continued perceptibly for months.

Another striking effect has been found to take place when a superconducting ring in which there is a current is allowed to rise to the temperature at which it regains its resistance. The current, which was, of course, a stream of electrons, suddenly stops, so that angular momentum of the stream about the axis of the ring is reduced to zero. But no external force was applied to the ring, and its angular momentum must continue. It is actually found by experiment that the ring does start forward perceptibly. This seems to prove that an electric current in a metal is really a stream of material particles of some kind.

Only a small number of metals become superconducting as they approach absolute zero. Most metals show no abrupt change of resistance near absolute zero, for instance, gold, silver, and copper, which can be obtained very pure. At present there is no accepted explanation of superconductivity.

THERMOELECTRICITY

464. Current Produced by Heat. Thermocouples.—If a circuit is formed of two different metals and a galvanometer and the junctions of the metals are brought to different temperatures, a current flows, which means that an e.m.f. is produced. This was discovered by Seebeck in 1826. The e.m.f. in volts for a copper-iron *thermocouple*, as it is called, is shown in Fig. 377. If one junction is kept at 0°C, and the other is heated the current is from copper to iron through the hot junction, and it increases until the hot junction reaches 275°C. After that it decreases, becomes zero at 550°C. and then reverses and continues in the reverse direction until the copper melts. A curve representing the relation is a parabola (Fig. 377). If one junction is kept at 100°C. while the other is heated, the result is also a parabola with the same axis, but the maximum e.m.f. is smaller.

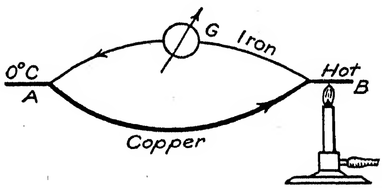


Fig. 376.

Copper-iron Thermo-couple (Cold Junction at 0°C.)

Temperature in °C. of Hot Junction.	E.m.f. in Volts.	Temperature in °C. of Hot Junction.	E.m.f. in Volts.
0	0.00	400	+0.00172
100	+0.00129	550	0.00
200	+0.00201	600	−0.00089
275	+0.00216		

Insertion of another metal in the circuit of Fig. 376 has no effect on the e.m.f., provided its ends are at the same temperature. Hence junctions can be soldered without producing any additional e.m.f., and a galvanometer may be included in the circuit if its junctions are at the same temperature.

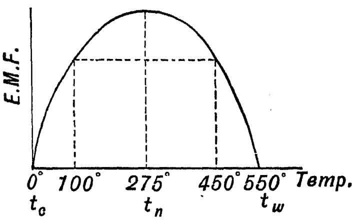


Fig. 377.

465. Converse Effect. Heating by Current.—If a current is sent through a thermocouple by a battery inserted in the circuit, heat is produced at one junction, which therefore becomes warmer, and heat is absorbed at the other junction, which is therefore cooled. This converse to Seebeck's principle was discovered by Peltier in 1834.

In one experiment he used a cross (Fig. 378) of antimony and bismuth. A current from the antimony *DC* to the bismuth *CB* cooled the junction *C*, and the galvanometer showed a deflection; reversing the current heated *C* and reversed the deflection. (The device is now used in *thermo-couple meters* for small currents.)

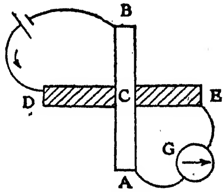


Fig. 378.

We can readily see, by the conservation of energy, which junction of a copper-iron thermocouple will be heated by the battery current. At one junction the current will go from copper to iron; and if it heated that junction, the heating would produce a thermoelectric current in the same direction as the battery current and this would add to the battery current and produce further heating at the junction and a greater thermoelectric current, and so on, the result being a perpetual motion machine. Hence the battery current must cool the junction at which it passes from copper to iron and heat the other junction.

466. Electron Explanation. Thomson Effect.—When two metals are in contact, the velocity and the density of the electrons in one are not the same as in the other; so there is a tendency for electrons to flow from one of the two metals to the other. In a circuit formed by joining a copper wire and an iron wire (Fig. 376), all at the same temperature and initially uncharged, electrons start to flow from the iron to the copper at each junction; but the process stops, almost at once, when the copper gets a certain oversupply of electrons. It begins again when one junction, *B*, is heated, for the

heating does not affect the velocity and the density of electrons in just the same way in the two metals. There are then different e.m.f.'s at the two junctions and a current is produced.

Apparently the effect should go on increasing as B is heated, and as it does not, something else must be happening. Reasoning by the principles of thermodynamics, long before electrons were thought of, led Thomson (later Lord Kelvin) to the same conclusion. He predicted from theory, and confirmed by experiment, that e.m.f.'s are also produced along the wires themselves, if there is a gradient of temperature. This means that parts of the same metal, if at different temperatures, act thermoelectrically as different metals. Electrons tend to flow from hot to cold in copper and in the opposite direction in iron, and so on for other metals, except lead in which there is no tendency in either direction.

It is now seen that the e.m.f. observed by Seebeck is really the resultant of four e.m.f.'s, two at the junctions and two Thomson e.m.f.'s along the wires themselves.

467. Thermoelectric Power.—For nearly all thermocouples, the curve of Fig. 377 is approximately a parabola over a wide range of temperature. We can therefore write for the e.m.f. E when one junction is kept at 0°C.

$$E = at + \tfrac{1}{2}bt^2$$

where a and b are constants for the thermocouple but vary with the purity and other conditions of the metals. For a certain copper-iron couple, E being in volts, a was 15.81×10^{-6} and b was -0.0576×10^{-6}.

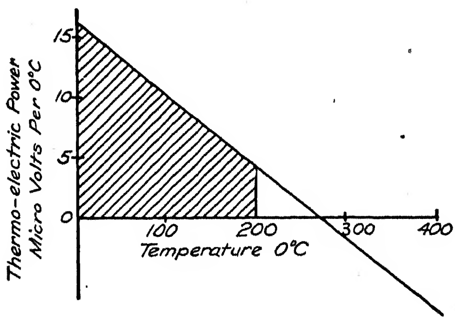


Fig. 379.

The change of E per degree change of temperature of the hot junction, that is dE/dt, is called the *thermoelectric power* of the couple. From the above

equation for E we get

$$\frac{dE}{dt} = a + bt$$

Thus the thermoelectric power is a linear function of the temperature. Since dE/dt is zero when E is a maximum, the temperature of maximum e.m.f. is $-a/b$, which for copper and iron gave 275°C. A plot of dE/dt is a straight line

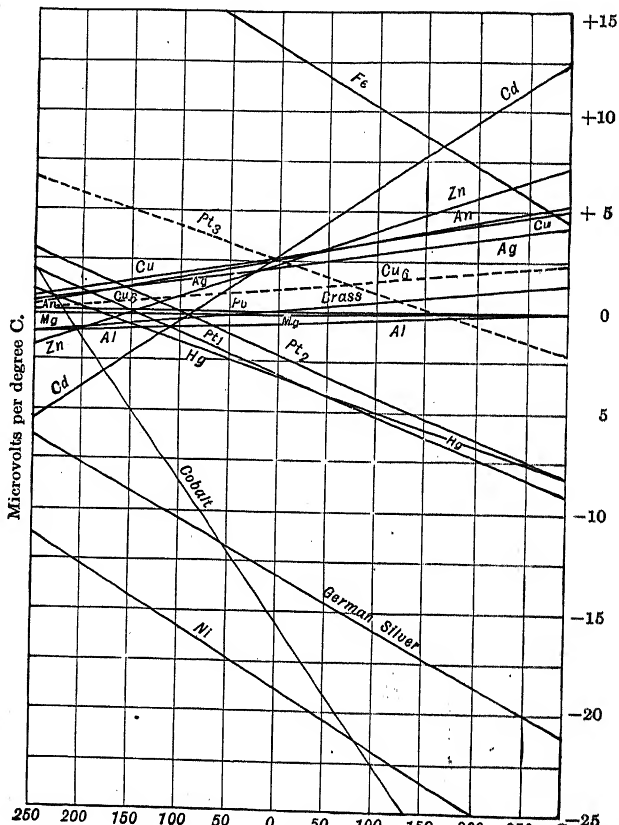


Fig. 380.

(Fig. 379) that cuts the t-axis at 275°C. This diagram represents the thermoelectric properties in a very simple way. For example, the e.m.f. when the junctions are at 0°C. and 200°C. is the average value of dE/dt in the interval multiplied by 200, and this is the area of the shaded part in the diagram.

In a thermoelectric diagram of different metals (Fig. 380) each metal is represented by a straight line with temperature as abscissa and thermoelectric power as ordinate and on the understanding that the other metal in the couple

is lead. Lead is chosen for this purpose because the Thomson effect in it is zero. The thermoelectric force for a couple of any two metals at given temperatures of the junctions, both temperatures being below the neutral point or both above it, is represented by the area bounded by the two lines and the two temperature ordinates, as will readily be inferred from Fig. 379. If, however, the two temperatures are on opposite sides of the neutral point, the area to be taken will be the difference of the areas of the two triangles with the neutral point as common vertex.

ELECTRIC CURRENTS IN LIQUIDS

468. Liquid Conductors.—Some liquids, such as mercury and molten metals, act just like solid metals in transmitting electric currents. A stream of electrons passes through them and the only effect on the substance of the conductor is due to heat generated. Liquids in this class need not be considered further here. But there are liquids which undergo chemical changes in addition to thermal effects, when a current passes through them, and these we shall now consider. Such liquids are solutions and the most important of them are aqueous solutions.

Water can be purified until its specific resistance is one hundred million times that of mercury. Ordinary distilled water is a much better conductor, and water from a faucet is a relatively good conductor, evidently owing to the presence in it of dissolved substances. All aqueous solutions of acids, bases and salts are fairly good conductors. There are, however, substances which do not form conducting solutions when dissolved in water. An aqueous solution of sugar is a non-conductor. Substances which form conducting solutions in water do not necessarily form conducting solutions in other solvents. An aqueous solution of common salt is a good conductor, but a solution of salt in alcohol is a non-conductor. After this brief statement about conduction in liquids in general, we shall now confine attention to conduction in aqueous solutions; and whenever the word solution is used alone, it will always mean an aqueous solution.

Conduction in a liquid accompanied by chemical changes is called *electrolysis*, and the liquid is called an *electrolyte*. The conductivity of electrolytes is due to the presence in them of charged atoms or groups of atoms, called *ions*. These are able to move in definite directions when the liquid is in an electric field.

469. Electrolytes.—A form of electrolytic cell that is convenient when gaseous products of electrolysis are to be collected is illustrated

in Fig. 381. Two platinum wires, A and K, called *electrodes*, are sealed in the glass and enter the liquid. If A is connected to the positive terminal of a battery it is called the *anode* (up road), and K, connected to the negative terminal, is called the *cathode* (down road). An ammeter measures the current.

When this apparatus is used for the electrolysis of a weak solution of hydrochloric acid, HCl, chlorine is released at the anode A and hydrogen at the cathode K, and the volumes released are equal. This net result is brought about by a number of operations in the cell.

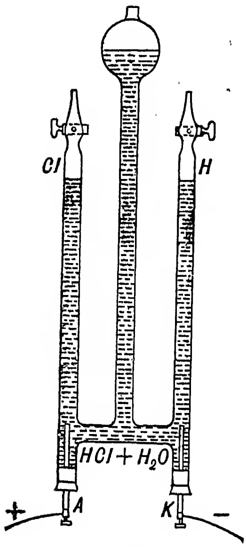


Fig. 381.

When HCl is dissolved in water, all of the HCl molecules break up into positive *ions*, or positively charged hydrogen atoms, represented by H^+, and negative ions, represented by Cl^-. This process is called *ionization*. H^+ is an atom that has lost one electron e^-, and Cl^- is an atom that has gained an electron.

When the electrodes are connected to the battery, an electric field is established between the anode and the cathode, and the Cl^- ions are drawn toward the anode and the H^+ ions toward the cathode. There they become normal atoms and molecules, the reactions being:

$$\underset{\text{(at anode)}}{2Cl^- \rightarrow Cl_2 + 2e^-} \qquad \underset{\text{(at cathode)}}{2H^+ + 2e^- \rightarrow H_2}$$

Thus chlorine gas is formed at the anode and hydrogen at the cathode.

The electric current in the solution is a stream of positive ions in the direction from the anode to the cathode and a stream of negative ions in the opposite direction. The current in the ammeter in the circuit is a stream of electrons from the anode, where electrons are being given up, to the cathode, where electrons are taken in (see the reactions above), although the (positive) current is, by definition, in the opposite direction.

Why dissociation of the HCl should take place, that is, why the H and Cl parts of the molecules should cease to attract each

other when the HCl molecules are in water, can be explained, in part at least, by a consideration of Coulomb's law of attraction between opposite charges (§368). Evidently the greater the dielectric constant of the medium, the less the attraction. Now the dielectric constant of water is extremely high, about 81, and it seems also to be true that other solvents in which dissociation takes place also have high dielectric constants; for example for alcohol, the next most common solvent, it is about 26.

470. Ohm's Law for Electrolytes.—Whether Ohm's law holds for an electrolyte can be tested by using batteries of different e.m.f.'s in succession in a circuit containing an electrolytic cell and an ammeter, or by some equivalent arrangement. It is thus found that the current I is not proportional to the e.m.f. E of the battery. It is, however, proportional to $E - E'$ where E' is a constant for the particular cell. This expression suggests at once that the electrolytic cell, when a current is passing through it, becomes itself a temporary battery, and that its e.m.f., E', is opposed to the e.m.f., E, of the battery. The supposition can readily be tested by connecting the cell to a battery for a time and then quickly replacing the battery by a galvanometer. A current that rapidly dies away flows through the galvanometer and the cell in the direction opposite to that of the original battery current. This confirms the supposition of an opposing e.m.f. set up in the cell. It is called a *polarization* e.m.f. Each electrolytic cell, consisting of electrodes of certain materials and a certain electrolyte, has a definite polarization e.m.f. of its own. If the e.m.f. applied to a cell is less than its polarization e.m.f., a current will start, but will not continue.

When polarization e.m.f.'s are taken into account, it is found that electrolytes obey Ohm's law fairly well in the form:

$$\frac{E - E'}{I} = R$$

There are, however, minor effects that should be taken into account, such as changes of concentration of the electrolyte at the electrodes, but we need not discuss them at present.

The cause of the polarization e.m.f. in an electrolytic cell is the presence, at the cathode, of a layer of H^+ ions waiting to take on electrons and, at the anode, of a layer of Cl^- ions waiting to discharge electrons. These two layers would by themselves produce an e.m.f. from the cathode to the anode in the liquid, and this e.m.f.

has to be subtracted from the primary e.m.f. to get the resultant e.m.f. It takes an appreciable time to build up these layers; and, when the primary e.m.f. is simply removed, the ions that form the layers drift back into the liquid.

Polarization e.m.f.'s are not produced in electrolytic cells when high-frequency alternating currents are used, for there is not time, while the current is running in one direction, for ion layers to be built up. When tested in this way, electrolytes conform to Ohm's law. Moreover, the resistance of an electrolytic cell can be measured by Wheatstone's bridge, with a source of alternating current substituted for the battery and a telephone receiver for the galvanometer.

Polarization e.m.f.'s are not set up by a current in an electrolytic cell if the electrodes are of the same material and their surfaces are not changed chemically by the current, for example when the electrodes are copper and the electrolyte is a solution of $CuSO_4$ (except for slight effects due to changes of concentration).

471. Faraday's Laws of Electrolysis.—Before the year 1833 the whole subject of the passage of electricity through liquids was in a vague, disconnected state. Then Faraday introduced the admirable set of terms we have been using (italicized in §469) and, from a carefully planned set of experiments, drew two fundamental conclusions, now known as *Faraday's laws of electrolysis:*

1. *The mass of an element liberated or deposited at an electrode is proportional to the quantity of electricity that passes.*

2. *The mass of an element liberated or deposited by the passage of a given quantity of electricity is proportional to the atomic weight of the element divided by its valence.*

We have already assumed the first law in sketching what happens in electrolysis, according to the modern view of the part played by electrons and their indestructibility, though, of course, electrons were not known in Faraday's time.

The first law can be stated as an equation. If a current i, flowing for a time t, liberates or deposits a mass m of an element

$$m = zit = zq$$

where z is a constant for the element called its *electrochemical equivalent.* If m is in grams and q in coulombs, z for an element is the number of grams of the element liberated or deposited by the passage of one coulomb.

Different ways of stating the second law have been found desirable. The atomic weight of an element divided by its valence is called the *chemical equivalent* of the element, valence meaning the number of electrons of the element that are effective in chemical combinations and in electrolysis. The chemical equivalent of an element is, then, the proportion in which it enters chemical combinations. The second law amounts to saying that the amount of a substance liberated in electrolysis is proportional to its chemical equivalent. It is evident from their definitions that the chemical equivalents of substances are proportional to their electrochemical equivalents.

It has been found convenient to have a term for the amount of an element in grams equal numerically to the atomic weight of the element. It is called the *gram-atomic-weight* of the element, which we may abbreviate to g.a.w. Now it has been found that 96,500 coulombs liberates one g.a.w. of hydrogen (atomic weight 1.008). On account of its frequent occurrence the quantity 96,500 coulombs has received a special name; it is called one *faraday*, which we may abbreviate to $1F$. With these terms we can state the second law numerically in two ways:

One F of electricity liberates or deposits 1 g.a.w. of any univalent element, $\frac{1}{2}$ g.a.w. of any bivalent element, $\frac{1}{3}$ g.a.w. of any trivalent element, and so on.

One g.a.w. of an element is liberated by $1F$ if the element is univalent, by $2F$ if it is bivalent, by $3F$ if it is trivalent, and so on.

Elements	Atomic weight	Valency	Chemical equivalent	Electro-chemical equivalent
Chlorine	35.46	1	35.46	.0003675
Copper	63.57	2	31.78	.0003294
Hydrogen	1.008	1	1.008	.00001046
Iron, ferrous	55.84	2	27.92	.0002893
Iron, ferric	55.84	3	18.61	.0001929
Oxygen	16.00	2	8.00	.00008291
Silver	107.88	1	107.88	.0011180
Zinc	65.37	2	32.68	.0003387

472. Interpretation of Faraday's Laws.—Faraday's laws, taken in connection with the atomic theory of matter, lead to the conclu-

sion that electricity is also atomic or consists of integral multiples of an individual unit; this seems to have been first pointed out by Helmholtz, before the modern electron theory was suggested by other phenomena.

One g.a.w. of *any* element contains the same number of atoms, and this number is known from various lines of evidence to be 6.04×10^{23}. One g.a.w. of an element is associated in electrolysis with $1F$, or $2F$, or $3F$, . . . , of electricity, depending on the valence of the element, but not with any fraction of one F. This seems to show that electricity, at least as it behaves in electrolysis, consists of indivisible units such that, if we denote the magnitude of the unit in coulombs, by e,

$$6.04 \times 10^{23} e = 96{,}500 \text{ coulombs}$$

and therefore

$$\begin{aligned} e &= 1.60 \times 10^{-19} \text{ coulomb} \\ &= 1.60 \times 10^{-20} \text{ e.m.u. of charge} \end{aligned}$$

The value just obtained for e is the same as that obtained for the electron by other methods, especially by Millikan's oil-drop experiment (§400). While an approximate value for e was found from electrolytic data before the modern electron theory (founded chiefly on conduction in gases) was accepted, in practice the calculation is now reversed. The charge of an electron and the charge called a faraday can be determined so much more accurately than the number of atoms in a gram-atomic-weight, that the last is calculated from the other two.

473. Products of Electrolysis.—In the electrolysis of hydrogen chloride solution (§469) the hydrogen and chlorine are *primary products* of the electrolysis. The change that takes place at the electrodes is simply a discharge of the ions of the electrolyte. In other cases the solvent or the electrodes may, by reactions with the primary products, give rise to *secondary* products. Some simple examples of this will illustrate the general principles of such reactions.

Consider the electrolysis of a *dilute* solution of sulphuric acid (H_2SO_4) in the apparatus of Fig. 381. The products are two volumes of hydrogen liberated at the cathode for every one volume of oxygen liberated at the anode. The net result is that the water is being decomposed rather than the sulphuric acid. In this case

the solvent, water, plays an important part. The purest water is slightly conducting (§468), which means that it is ionized to some slight extent. There are present in the solution H^+ and $SO_4^=$ ions from the H_2SO_4, and H^+ and OH^- ions from the water. At the cathode the H^+ ions are discharged, but at the anode it is the OH^- ions that are discharged, rather than the $SO_4^=$ ions, the former giving up their extra electrons more readily than the latter. The reaction at the anode is

$$4OH^- \rightarrow 2H_2O + O_2 + 4e^-$$

These four electrons, produced with each molecule of oxygen, flow into the outer circuit at the anode, and from its other end at the cathode four electrons issue and discharge four hydrogen ions at the cathode, producing two molecules of hydrogen. Though the sulphuric acid does not play any part in the reactions at the electrodes, the ions formed from it conduct the current through the solution.

In the electrolysis of a dilute Na_2SO_4 solution hydrogen and oxygen are evolved at the electrodes, as of the preceding case and in the same proportions as before. In the solution there are Na^+ and $SO_4^=$ ions, together with H^+ and OH^- ions. The OH^- ions are discharged more readily than the $SO_4^=$ ions, and the H^+ ions take on electrons more readily than the Na^+ ions do.

Electromotive Series of the Metals (Referred to Hydrogen)

Metal	Symbol	Normal Electrode Potential
Sodium	Na^+	−2.71
Magnesium	Mg^{++}	−2.4
Aluminum	Al^{+++}	−1.7
Zinc	Zn^{++}	−0.76
Iron	Fe^{++}	−0.44
Cadmium	Cd	−0.40
Tin	Sn^{++}	−0.13
Lead	Pb^{++}	−0.12
Hydrogen	H^+	0.00
Copper	Cu^{++}	+0.34
Mercury	Hg^+	+0.79
Silver	Ag^+	+0.80

As regards the greater or less readiness with which the atoms of different metals give up electrons and become ions or the converse, all metals can be arranged in a definite series from which it can be seen what will happen in any particular case. In the Table a number of metals are arranged so that the atoms of a metal nearer the *top* have a greater tendency than those of one farther down to give up electrons and become positive ions. Now atoms that are more readily ionized are more persistent in remaining ionized. So an ion of a metal nearer the top of the list has less tendency to become neutral, and therefore less tendency to be liberated or deposited in electrolysis, than an ion of a metal lower in the list. The figures in the third column express the same thing numerically, as will be explained later (§477).

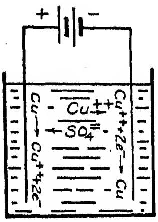


FIG. 382.

474. Effects of Electrodes on Products of Electrolysis.—It will be instructive to compare the electrolysis of a dilute solution of copper sulphate, first with platinum electrodes and next with copper electrodes. In both cases dissociation of the copper sulphate gives Cu^{++} and $SO_4^{=}$ ions, and that of the water gives H^+ and OH^- ions. What takes place at the cathode is the same in the two cases. Both Cu^{++} ions and H^+ ions are attracted to the cathode, but the former take on electrons more readily than the latter do. The result is represented by the reaction:

$$Cu^{++} + 2e^- \rightarrow Cu$$

The copper produced is deposited on the cathode, whether it is of platinum or copper, and two electrons per atom of copper deposited are taken in from the cathode, to which they are being driven by the battery through the external part of the circuit.

At the anode, when it is of platinum, oxygen is liberated. The process is the same as in the electrolysis of H_2SO_4 (§473), the OH^- ions being discharged more readily than the $SO_4^{=}$ ions, so that water and oxygen are produced and electrons go to the anode, to be conveyed to the cathode by the battery. But when the anode is of copper, no oxygen is produced. Atoms from the anode go into solution as cupric ions, Cu^{++}:

$$Cu \rightarrow Cu^{++} + 2e^-$$

The cupric ions travel through the liquid to the cathode and the electrons go through the battery to the cathode. The concentration of copper sulphate in the solution does not change, the $SO_4^{=}$ taking no part in the reactions.

Metals in general, when used as anodes in electrolytic cells, act like copper in producing positive ions; but platinum and gold are exceptions. At the cathode either hydrogen or the metal may be liberated or deposited.

475. Electricity Produced by Chemical Action.—The discovery, about one hundred and fifty years ago, that electricity could be produced by chemical action, was made in a curious way. Galvani of Bologna noticed that a leg of a freshly killed frog twitched when two metals, iron and brass, touched the leg and also touched each other. Galvani recognized this as an electrical effect, for similar effects had been produced by an electrostatic machine. Though Galvani studied this curious effect for several years, it was Volta of Pavia who (in 1799) developed it into the voltaic cell, a device for transforming chemical energy into the energy of an electric current. Many of the fundamental discoveries in Electricity were made with voltaic cells. Many such cells have been invented and these primary cells, as they are called, have not yet outlived their usefulness. The following brief account of the subject will be confined to the fundamental principles involved in this method of producing an electric current.

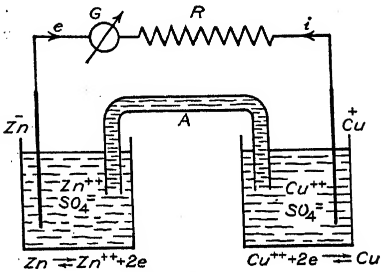


FIG. 383.

476. Principle of a Primary Cell.—If a rod of zinc is dipped into a dilute solution of zinc sulphate, some of the zinc atoms go into the solution as zinc ions, Zn^{++}, the electrons being retained by the zinc and making it negatively charged; but the process is soon arrested by the accumulation of the two charges. If, in another vessel, there is a copper rod in a copper sulphate solution, copper atoms go into solution as copper ions, Cu^{++}. If the two rods are now connected by a wire R (Fig. 383 without A), electrons flow from the zinc to the copper through R, because zinc has a much greater tendency than copper to lose electrons and become ionized (Electro-

motive series, §473). The supply of electrons to the copper plate reverses the action there by neutralizing Cu^{++} ions, so that copper is deposited on the copper plate leaving an excess of $SO_4^{=}$ ions. With the accumulation of an excess of Zn^{++} ions in one vessel and of $SO_4^{=}$ ions in the other, the flow of electrons in R is soon stopped.

What is needed to enable the flow to resume and continue is a connection between the vessels, so that the excess Zn^{++} ions and the excess $SO_4^{=}$ ions can flow together and neutralize, forming $ZnSO_4$, without allowing the liquids as a whole to mix. A bent glass tube A containing some electrolyte will act as a "bridge" for the purpose, but we shall see that there are other ways of accomplishing it. The copper is called the positive *pole* because the current, which is opposite in direction to the stream of electrons, is from the copper to the zinc in the wire.

The general plan of such a cell is to get a stream of electrons in a wire by making use of the tendency that a substance high in the electromotive series has to give off electrons and form positive ions, and then to neutralize these positive ions by negative ions from a metal lower down in the series.

If we now suppose the connecting wire R to be removed, the difference of potential of the poles will be the e.m.f. of the cell. This will depend on the concentrations of the solutions. What it would be if both solutions were normal solutions is found from the table of normal electrode potentials (§473) to be $0.34 - (-0.76)$ or 1.10 volts.

477. Concentration of Solutions. Electrode Potentials.—The starting point in the action of the cell we have been considering seems to be the flow of positive ions from the zinc electrode into the solution. But there is also some return flow, and this is greater the greater the number of zinc ions already in the solution. The current depends on the excess of zinc ions given to the solution over the number received from it. Thus the current carried by the zinc ions is greater, the weaker the solution of zinc sulphate around the zinc electrode. On the other hand, to neutralize the stream of zinc ions calls for a large supply of negative $SO_4^{=}$ ions from the copper sulphate. Thus the e.m.f. of a cell depends on the nature and concentration of the solutions, as well as on the nature of the electrodes. For the cell we have been considering, the e.m.f. is a maximum when the zinc chloride solution is dilute and the copper sulphate solution is concentrated.

Cells that will give currents can be formed with electrodes of the same metal, say zinc, and solutions of the same salt, say zinc chloride, provided the concentrations of the solutions around the two electrodes are unequal. Such cells are called *concentration cells.* Their e.m.f.'s are small fractions of a volt.

The term "normal electrode potential" at the head of the third column of the Table of §473 may now be explained. A "normal" solution is one that contains, per liter of the solution, an amount of the dissolved substance equal numerically in grams to the molecular weight of the substance divided by the valence of the metallic ion. For example, for zinc chloride this is $(65.38 + 2 \times 35.45)/2$ or 68.14 grams. But the electrode potentials in the table are not absolute potentials. In reality, only differences of potential in the series are of practical use. So any zero level of potential may be assumed (compare temperature). The zero chosen arbitrarily is specified as the electrode potential of a hydrogen electrode in a normal solution of an acid (*i.e.* a hydrogen salt).

478. The Daniell Cell.—Replacing the salt "bridge" in the cell of Fig. 383 by a cup of unglazed earthenware turns the cell into a Daniell cell. The porous cup stands in a glass vessel that contains the copper electrode and the concentrated copper sulphate solution, and in the cup is the zinc electrode and some solution. Different solutions surrounding the zinc electrode have been used, such as zinc sulphate, sulphuric acid, and sodium chloride. The only essential thing, as regards producing a current, is that zinc atoms from the zinc electrode go into solution as zinc ions, leaving the corresponding electrons behind them on the electrode.

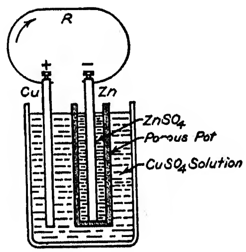


Fig. 384.

When the cell is in operation the zinc electrode is eaten away and copper is deposited on the copper electrode. Stated chemically, the net result is the formation of zinc sulphate and the decomposition of copper sulphate. From the known energy of these reactions the e.m.f. of the cell can be calculated. Suppose that one coulomb of electricity passes through the cell. From the table of electrochemical equivalents (§471) we can calculate how much zinc goes into solution and how much copper is separated out. From these results and a table of heats of formation of compounds we can

readily calculate the energy supplied by each of the two reactions. The result is found to be 1.09 joules. In the present case, the energy is not developed in the form of heat, but as the energy of an electric current. If the e.m.f. of the cell is E, the work done by the passage of q coulombs is Eq joules (§422), and, putting $q = 1$, we get

$$E \text{ (volts)} \times 1 \text{ (coulomb)} = 1.09 \text{ (joules)}$$

This gives 1.09 volts as the e.m.f. of the cell, which is very close to the observed value. It is practically independent of the temperature of the cell (if it were not the calculation would be incomplete).

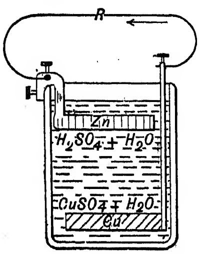


Fig. 385.

The *gravity cell* (Fig. 385) is a rough form of the Daniell cell. It dispenses with the porous cup and depends on gravity to keep the solutions separated. It is very readily set up. The zinc electrode hangs from the side of the jar at the top, and connection is made with the copper electrode, which is at the bottom of the jar, by an insulated wire. The jar is filled with a solution of sulphuric acid; and copper sulphate crystals, dropped to the bottom, produce a dense solution of copper sulphate. The e.m.f., about a volt, varies with the state of the cell, but the internal resistance is low and the cell can give a considerable current.

479. Weston Standard Cell.—A reliable standard of e.m.f. is needed for the same reasons, scientific and legal, that make a reliable standard of length or mass necessary. Hence a great deal of research has been devoted to the design of a special cell that is not intended to give any appreciable current but to serve as a standard of e.m.f.

We have seen that the Daniell cell has a rather definite e.m.f. and practically zero temperature coefficient, two very desirable qualities for a standard. Replacement of the copper and copper sulphate by mercury and mercurous sulphate gave a better standard (the Clark cell). Further modification, amounting in principle to the additional replacement of zinc and zinc sulphate by cadmium and cadmium sulphate, gave a still better standard, namely the Weston cell (Fig. 386), which, when prepared in a carefully specified way, is now the legal standard of e.m.f. In actual structure the cell is rather complex. It is made of a glass tube of H-form, with platinum wires sealed in the glass to make contact with the elec-

trodes of mercury and cadmium amalgam. Above the mercury is a paste of mercurous sulphate and cadmium sulphate and above that are cadmium sulphate crystals. Above the cadmium amalgam in the other arm are cadmium sulphate crystals. Filling most of the rest of the tube is a solution of cadmium sulphate. The mercury is the positive pole of the cell, as might be inferred from the electromotive series (§473).

When prepared according to specifications, the e.m.f. of the Weston cell is 1.0183 volts at 20°C. with a temperature coefficient of -0.00004. It is used in potentiometer work (§461) in series with such a high resistance that the current through it is extremely minute; even a few millionths of an ampere would seriously damage it as a standard of e.m.f.

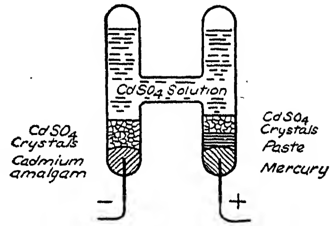


FIG. 386.

480. The Dry Cell.—A cell, called the Leclanché cell, that was much used at one time, had electrodes of zinc and carbon with ammonium chloride (NH_4Cl) as electrolyte. This was developed later into the "dry" cell, in which there is no spillable liquid. An outer zinc can serves as negative electrode. Inside of it as a lining is absorbent pulp-board, impregnated with a paste of flour and starch that is wetted by ammonium chloride solution. Within the lining is a mixture of manganese dioxide (MnO_2) and carbon, and a central carbon rod serves as positive electrode.

The zinc serves its usual purpose of supplying zinc ions, and in the reaction of these with the ammonium chloride hydrogen ions are produced. The function of the manganese dioxide is to supply oxygen to these hydrogen ions and produce water, with the aid of electrons that come around from the zinc through the external circuit to the carbon electrode. The stream of electrons from zinc to carbon constitutes the real current in the external circuit; but, as usual, the current is regarded as being in the opposite direction. This is, however, a somewhat simplified account of the reactions that take place in the cell.

481. Lead Storage Cell.—A primary cell, such as we have been considering, converts chemical energy into the energy of an electric current. A *secondary* or *storage cell* is one that has to be charged with chemical energy by the action of a current, before it will give a current.

The most common form of storage cell, called the *lead cell*, was invented by Planté (1859). Lead plates were used as electrodes in an electrolytic cell containing a sulphuric acid solution. Currents were sent through the cell, alternately in opposite directions, for

months, until the surfaces of both plates were in a spongy condition, one being lead (Pb) and the other lead peroxide (PbO_2). Such a cell gave a discharge current opposite in direction to the last current applied to it, and the surfaces of the plates became covered with lead sulphate ($PbSO_4$). The tedious process of formation was greatly reduced by Faure (1881) by starting with plates already pasted with lead and lead peroxide. In course of use these surfaces tended to fall off. In modern cells lead grids are used as a framework and are filled with pastes consisting of different mixtures of lead, oxides of lead, and sulphuric acid that have been found suitable. But, for the best cells, the positive PbO_2 plate, the one *from* which the current issues in discharge, is still "formed" by the Planté method.

In discharging, the storage cell acts much like a primary cell. In the sulphuric acid there are H^+ and $So_4^=$ ions. The current leaves by the PbO_2 plate and the H^+ ions go to it. There they reduce PbO_2 to PbO, forming water, and then the PbO and H_2SO_4 form $PbSO_4$ and more water. In the charging of the cell these actions are reversed, the current entering by the PbO_2 plate. The net results may also be summarized thus:

$$\underset{\text{pos. pl.}}{PbSO_4} + 2H_2O + \underset{\text{neg. pl.}}{PbSO_4} \underset{\text{discharge}}{\overset{\text{charge}}{\rightleftarrows}} \underset{\text{pos. pl.}}{PbO_2} + 2H_2SO_4 + \underset{\text{neg. pl.}}{Pb}$$

In charging, the positive plate, originally $PbSO_4$, contains lead in the divalent state Pb^{++}, and during the charging process this is transformed into PbO_2 or quadrivalent lead ions Pb^{++++}. The negative plate, on the other hand, takes on electrons, so that the Pb^{++} lead ions become neutral lead atoms. On discharge, the reverse of this process takes place. The charging produces a decrease of water and an increase of sulphuric acid, so that the density of the liquid is increased. The opposite takes place in discharge. Thus the density of the solution, as found by a hydrometer, may be used as a test of the state of charge of the cell.

The e.m.f. of a lead cell, when fully charged, is about 2.1 volts, but it falls fairly rapidly on discharge to about 2.0 volts, where it remains, almost constant, until the discharge is nearly complete, when it drops rapidly to about 1.8 volts. The resistance is low, of the order of 0.01 ohm for a cell of ordinary size. The cell is there-

fore especially useful where a large current is required for a short time, as in automobile starters.

482. The Edison Storage Cell.—In this form of storage cell the positive plate is nickel peroxide (Ni_2O_3), packed in a steel grid with flaked nickel to make good contact with the grid. The negative plate is iron, and the electrolyte is a solution of potassium hydroxide (KOH). The reactions in charge and discharge are essentially:

$$\underset{\text{pos. pl.}}{Ni_2O_3 + 3H_2O} + \underset{\text{neg. pl.}}{Fe} \underset{\text{charge}}{\overset{\text{discharge}}{\rightleftarrows}} \underset{\text{pos. pl.}}{2Ni(OH)_2} + \underset{\text{neg. pl}}{Fe(OH)_2}$$

The cell has a lower e.m.f., about 1.4 volts, and a higher resistance than a lead cell, so that it does not give as great a current. But, for a given energy content, it is lighter than a lead cell, and it can stand much more abuse electrically and mechanically; for example, it is not injured by being left uncharged. It is therefore especially useful for vehicles, the purpose for which it was originally designed.

ELECTROMAGNETIC INDUCTION

483. Induction of a Current by a Magnet.—Up to 1831 three methods for producing an e.m.f. were known: (*a*) by an electrostatic machine, (*b*) by a thermoelectric couple, (*c*) by a voltaic cell. Then Faraday added a fourth, namely (*d*) by electromagnetic induction. Few discoveries in science have had such important results.

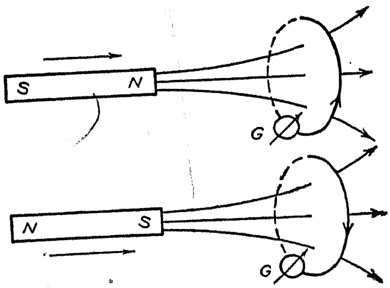


FIG. 387.

If the *N* pole of a magnet is brought up to a fixed coil of wire connected to a galvanometer (Fig. 387), a momentary current is produced in the coil. If the magnet is withdrawn a momentary current flows in the opposite direction. Bringing a *S* pole up to the coil starts a momentary current opposite in direction to the first current, and removing it starts a current in the direction of the first current. Increasing the speed of the motion of the magnet increases the maximum strength to which the current rises, but it does not affect the total amount of electricity that flows through a section of the wire. While we have spoken of the magnet as being moved, similar results are obtained if the magnet is at rest and the coil is moved. It is the relative motion of the two that gives rise to the current. This is called *electromagnetic induction*.

The current in the wire is due to some condition at all parts of the coil that must be regarded as an e.m.f. acting on the electrons in the wire. It must evidently be due to the magnetic flux from the magnet passing through the coil. It lasts only while the relative motion continues and is greater the greater the speed of relative motion. It is called an *induced e.m.f.*

In Fig. 387(a) the current in the coil, as seen in the direction of the axis of the approaching magnet, is counter-clockwise, that is, opposite to the amperian currents in the magnet or to a current in a solenoid equivalent to the magnet. We therefore call this a *negative* current. The coil with the current in it is equivalent to a magnet with its N pole turned toward the N pole of the approaching magnet. There is therefore a repulsion between the two, due to their magnetic field being opposed. Withdrawing the magnet causes a clockwise or *positive* current and a reversal of the polarity of the coil, so that now the fields are in the same direction and there is attraction between coil and magnet. It is readily seen that, in the other cases also, *the motion of the magnet is opposed by the magnetic field of the current.* We shall return to this point later (§485).

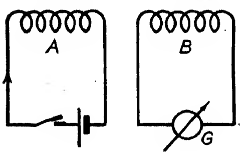


Fig. 388.

484. Induction of a Current by a Current. Faraday also found that a current could be induced in a coil, not by the use of a magnet, but by means of a current in another coil. The second coil could be used, like the magnet, to induce currents when moved towards or away from the first coil, so as to increase and decrease the magnetic flux through the first coil. But, with a current replacing the magnet, motion was not necessary. When a current is started in a stationary coil A (Fig. 388) by pressing a key, a momentary current is set up in a coil B that is parallel to A. This induced current is in the opposite direction to the current in A, so that the magnetic fields of the two coils are opposed, and the current in B is therefore a *negative* current. When the key is released and the current in A stops, there is a momentary *positive* current induced in B. All these effects are still more marked if one coil surrounds the other (Fig. 389).

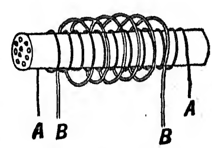


Fig. 389.

Faraday also found that a rod of soft iron in the primary coil A greatly increased the currents induced in the secondary, B, at make or break of the primary current. The presence of the iron increased the magnetic flux in the primary and in the field in which the secondary was. With this arrangement, making and breaking of the primary were not necessary. If, when the primary current was flowing, the iron rod was thrust into the primary, the increase of flux produced a secondary current. The effect of the presence of iron is still greater if both primary and secondary are wound on a soft iron ring (Fig. 390). Then practically the whole of the magnetic flux from the primary passes through the secondary, and the inductive action is very great. This, with the substitution of alternating currents in the primary producing alternating currents in the secondary, is essentially the alternating current transformer.

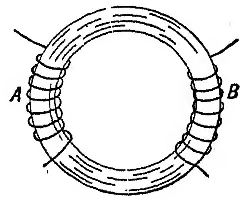


Fig. 390.

485. Work and Energy in Induction of Currents.—We have seen that, whenever a current is induced in a coil by the motion of a magnet, the current, acting through its magnetic field, exerts a force on the magnet that, in all cases, opposes the motion of the magnet. If the coil is moved and the magnet stays at rest, there will be a force opposing the motion of the coil. This inference from Faraday's experiments was drawn by Lenz in 1834, before the principle of the conservation of energy had been established; but we can see now that it follows directly from that principle. For if the induced current exerted a force that tended to increase the motion, it would thereby tend to increase the induced current itself, with a further increase of the force, and so on. Thus an unlimited amount of energy could be created, which would contradict the conservation of energy. Or we may say that the energy of the current could not be produced without doing work against some opposing force.

In the case of the induction of a current by a current a similar principle applies. When the coils are at rest, making the primary current produces a negative secondary current, that is, a current whose magnetic field is opposite to that of the primary and opposes the growth of the primary current. At break of the primary current the induced secondary current is positive and tends to prolong the

primary current; in fact it shows its action by a spark at the primary spark gap or key.

All of these relations are summarized in a very useful way by what is now known as LENZ'S LAW: *An induced current is in such a direction as to oppose the action that induces it.* This often provides a handy short cut in finding the direction of an induced current.

486. Expression for Induced E.M.F.—The *magnitude* of an e.m.f. induced by a changing magnetic flux through a circuit is given by what is now known as FARADAY'S RULE: *The induced e.m.f. is proportional to the time-rate of change of magnetic flux through the circuit.* If ϕ_1 is the magnetic flux through the circuit at any time and ϕ_2 that at a time t seconds later, the average rate of increase of flux in the time t is $(\phi_2 - \phi_1)/t$. This, according to Faraday's rule, is proportional to the average e.m.f., and, by a proper choice of units, we can always turn a proportionality into an equality. So we can write the expression for the average e.m.f. E_a induced in the time t thus:

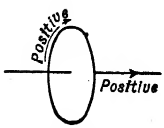


FIG. 391.

$$E_a = -\frac{\phi_2 - \phi_1}{t}$$

The reason for the minus sign can readily be seen. If we take, as usual, positive direction around a circuit and positive direction through it in accordance with the corkscrew rule (§424), then a positive $(\phi_2 - \phi_1)/t$ produces a negative E_a (Lenz's law).

The above expression for E_a is an average over time t, and it is sufficient if the flux is changing at a uniform rate. If, however, the rate is not uniform, it is necessary to consider what E is at any moment. The relation is just like that between the mean speed of a train between stations and its speed at any moment. The instantaneous rate of increase of flux means the value that $(\phi_2 - \phi_1)/t$ approaches as t is decreased toward zero. This is expressed, in the very convenient calculus notation, by $d\phi/dt$. Hence if the instantaneous value of the e.m.f. is E,

$$E = -\frac{d\phi}{dt}$$

As regards the units in these equations, we shall see a little later (§489) that, if ϕ is in maxwells (§440) and t in seconds, E is in electromagnetic units. If we now wish to change the equations so as to

give E in volts, we must divide the right-hand sides by the number of e.m.u. in a volt, which is 10^8.

These equations for E_a and E are for a circuit of a single turn. For a circuit of n turns, in each of which the same e.m.f. is induced, the right-hand sides must be multiplied by n.

It is frequently convenient to have E expressed in terms of changing field strength. If the field strength is H in oersteds and the permeability μ of the medium is uniform, $\phi = \mu HA$, where A is the area of a coil, supposed to be plane. Hence in this case

$$E_a = -\mu A \frac{H_2 - H_1}{t} \qquad \text{and} \qquad E = -\mu A \frac{dH}{dt}$$

Here also to get E_a or E in volts the right-hand sides must be divided by 10^8.

487. Induced Current and Charge.—From the formula for an induced e.m.f. and Ohm's law we can get expressions for the induced current and for the total flow through a section of the conducting wire. For simplicity let us confine ourselves to the magnitudes of the e.m.f., current and quantity and drop the minus signs. Consider first the case of a single turn of wire that forms part of a circuit of resistance R e.m.u. and suppose that the magnetic flux through it changes by $(\phi_2 - \phi_1)$ maxwells in t seconds. From the expression we have just obtained for the induced e.m.f. we get, by applying Ohm's law,

$$I_a = \frac{\phi_2 - \phi_1}{Rt} \qquad \text{or} \qquad I = \frac{1}{R}\frac{d\phi}{dt}$$

The induced charge Q e.m.u. that flows through each cross-section of the conductors that form the circuit is It or the sum of Idt. Hence

$$Q = \frac{\phi_2 - \phi_1}{R}$$

For a coil of n turns these results are to be multiplied by n.

The corresponding equations in practical units for a coil of n turns are now readily seen to be (omitting minus signs)

$$E = n\frac{d\phi}{dt} \times 10^{-8} \qquad I = \frac{n}{R}\frac{d\phi}{dt} \times 10^{-8} \qquad Q = n\frac{\phi_2 - \phi_1}{R} \times 10^{-8}$$

The induced charge Q can be measured by a ballistic galvanometer in the circuit. This is therefore a method of measuring a

change of flux $(\phi_2 - \phi_1)$, since R is readily found. The expression for Q does not depend on t, the time in which the change $(\phi_2 - \phi_1)$ takes place; but, for accurate measurements by a ballistic galvanometer, the time must be short, so that the charge shall have passed before the galvanometer has moved appreciably.

488. Induction in Space.—It is natural to enquire what, if anything, happens in experiments like those illustrated by Fig. 387 if there is no conducting circuit present or if the space is a vacuum. The answer we give is that everything we know about electromagnetic phenomena agrees with the belief that *closed electric lines of force are set up in any region where there is a changing magnetic flux.*

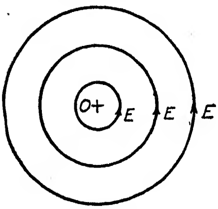


FIG. 392.

Fig. 392 shows some of the closed electric lines of force set up when a magnet, held at right angles to the paper with its N pole above O, is moved toward the paper. If a metallic circle coincided with one of these circles, a current would flow in the direction of the arrow (a flow of electrons in the opposite direction). The figure would represent equally well the closed magnetic lines of force surrounding a wire through O and perpendicular to the paper, if a current upward from the paper flows in it.

This is a good illustration of the reciprocal relation of electricity and magnetism: charges in motion have magnetic fields, magnetic poles in motion have electric fields.

489. Induction by "Cutting" Magnetic Lines.—It is often necessary to find the e.m.f. induced in a circuit when the change of flux through it is due to the motion of a single conductor that forms part of the circuit. For example, $PQRS$ (Fig. 393) is a circuit that is perpendicular to a magnetic field, and PQ can slide along SS' and RR', thus "cutting" the lines of magnetic flux. To find the e.m.f. E induced in the circuit $PQRS$, let us consider the question from the point of view of the conservation of energy. For simplicity we shall use electromagnetic units and denote the magnitudes of the e.m.f. and the current, without regard to their directions, by E and I.

If PQ is moving away from RS at a constant speed v, the magnetic flux through $PQRS$ is changing at a constant rate, and there is a constant current I in PQ, as well as in the rest of the circuit. PQ is then subject to a force of magnitude BIl (§433), where l is

the length of PQ, and, by Lenz's law (§485), this force is towards RS. Hence to keep PQ moving requires the application of an opposite force, say F, equal in magnitude to BIl. In time t, PQ moves a distance $d = vt$ and the work done by F is Fd or $BIlvt$. This is work done by some external agent in producing the current I.

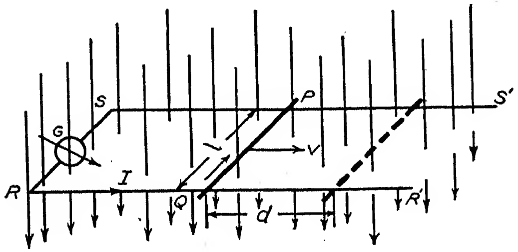


FIG. 393.

Now consider the work done by the current. The flow of electricity in the time t is It. Hence the work done by the current in time t is EIt. If the circuit is wholly a wire circuit, this work produces heat; but it might be used to produce chemical energy or mechanical energy. However it is spent, it must, by the conservation of energy, be equal to the work done by the force F. Hence

$$EIt = BIlvt$$

and therefore

$$E = Blv$$

Since lv is the rate of increase of the area of the circuit, Blv is the rate of increase of flux through the circuit, B being in gausses and Blv in maxwells per second. Thus Blv is a special form, suited to the present case, of the more general expression $(\phi_2 - \phi_1)/t$ for rate of change of flux.

If the flux through the coil is not at right angles to it but makes an angle θ with it, then the e.m.f. in volts is $E = \frac{Blv}{10^8} \sin \theta$, since only the component of the field perpendicular to the coil is effective in producing an e.m.f.

In this way we have arrived at the formula for Faraday's law by applying the conservation of energy. Moreover, we have found that the constant of proportionality in the expression for Faraday's law is 1, as assumed (§486).

To find whether the current in PQ is from P to Q, or from Q to P, note that its magnetic lines of force are circles around it, their direction being given by the corkscrew rule (§486). Where they pass through $PQRS$ they must be opposed to the field flux (§485). Hence if PQ is moving away from RS, the current in PQ is from Q to P.

These relative directions are also summarized in the *right hand rule* or *dynamo rule: If the right hand be held so that the forefinger is in the direction of the flux, the thumb in the direction of the motion, the center finger will be in the direction of the current.* This should be compared with the similar *motor rule* or left hand rule in §433.

490. Electron Explanation of Induction.—Consider again the arrangement of Fig. 393. In PQ there are numerous free electrons, negatively charged, and easily set in motion to form a current. There are also positively charged atoms and molecules, but we may treat them as being stationary in the wire. Now when PQ is moving away from RS the free electrons in PQ are being carried along in the direction in which PQ is moving. They therefore constitute a convection current (§416). This convection current of electrons is perpendicular to a magnetic field. To find in what direction the magnetic field impels it we must first, as usual, suppose it replaced by a positive current in the opposite direction, that is from PQ to RS. We then find by the left hand rule (§433) that the positive current is subject to a force in the direction P to Q, that is, P is at a higher potential than Q and the current around the circuit is therefore in the direction $PSRQ$.

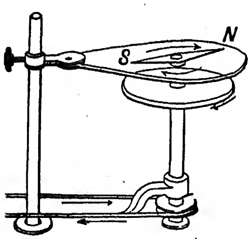


Fig. 394.

491. Electromagnetic Reactions.—Induced currents are due to the action of some agent, such as a magnet or another current, and they exert reactions on the agent that produces them. An example, known before Faraday's work, is illustrated by Fig. 394. A magnetic needle pivoted over a rotating copper disk rotates with the disk but at a slightly lower rate, so that there is some "slip," as it is called, between disk and needle. The rotation of the disk in the field of the magnet induces currents in the disk and these, by Lenz's law, oppose relative motion of disk and needle, so that the needle tends to follow the disk but at a slightly lower rate.

Another striking example is the damping of the motion of a copper disk that swings, like a pendulum, between the poles of a powerful magnet (Fig. 395). Here again induced currents in the disk arrest the relative motion that starts them. As the magnet is too heavy to be set in motion, the disk stops. If there are slots in the disk, the damping is greatly reduced, because the paths that the currents would naturally take are cut. A disk of aluminum or iron is only slightly damped, owing to its higher electrical resistance.

The vibrations of the rotating coils of galvanometers are usually damped by currents induced in a copper frame on which the coil is wound, and similar methods are used for arresting the index needles of ammeters and voltmeters.

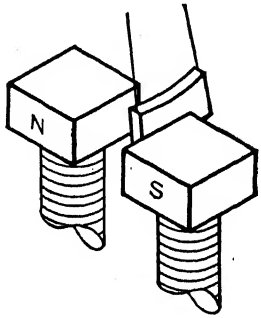


Fig. 395.

492. Self-induction.—In the experiment of Fig. 389 an increasing current in the primary produced an increasing magnetic flux through the secondary and gave rise to a negative e.m.f. in the secondary. Now the same increasing flux also passed through the primary and must have produced a negative e.m.f. in the primary itself. All this is true if there is no secondary, so let us suppose it absent. We now see that a changing current in a coil, accompanied by a changing magnetic flux through the coil due to the current, gives rise to an induced e.m.f. in the coil. This is called "self-induction." The direction of the induced e.m.f. in each case is given by Lenz's law: *it always opposes the change of current.* Hence growth of a current is retarded by self-induction, and decrease of a current is also retarded by self-induction. A current cannot be started or stopped instantaneously. This is the property, inertia, that we associate with ordinary matter, and thus we may consider self-induction as due to the inertia of the magnetic field.

The self-induction of a coil can be greatly reduced, without reducing its resistance, by winding it back on itself, so that opposite fluxes neutralize, and the coil is inductionless. The coils of resistance boxes are usually wound in this way (Fig. 365), but self-induction cannot be entirely eliminated, for two windings, especially insulated windings, must be separated to some extent.

493. Self-inductance. Mutual Inductance.—To find an expression for the e.m.f. of self-induction in a coil (or a circuit) we must

begin with the magnetic flux ϕ in the coil. If the flux density is B and the area of the coil A, and if the permeability μ of the medium is constant, the flux density B is μ times the field strength H and

$$\phi = BA = \mu HA$$

For a coil of N turns the total flux linking the N turns or *the flux turns* is $N\phi$, and for a medium of constant permeability ϕ is proportional to the current I. Hence

$$N\phi = LI \qquad \text{or} \qquad L = \frac{N\phi}{I}$$

where L is a constant for the coil, called its *self-inductance* or *coefficient of self-induction.* The self-inductance of a coil is *the total flux linking the coil due to unit current in it.* The self-inductance of a circuit is defined in a similar way.

The magnitude of the self-induced e.m.f. in a coil when the flux through it is changing is equal to the time rate of change of $N\phi$. If we take, as usual, the positive directions through and around a coil according to the corkscrew rule, the induced e.m.f. is negative when the flux is increasing. Expressing rate of increase of flux by $d\phi/dt$ and rate of increase of current by dI/dt, we can now write

$$E = -N\frac{d\phi}{dt} = -L\frac{dI}{dt}$$

Hence we can define self-inductance in another, but equivalent, way: *The self-inductance of a coil is the magnitude of the e.m.f. induced in it by unit rate of change of the current in it.*

From this we can define the electromagnetic unit of self-inductance in terms of the e.m. units of e.m.f. and current. The practical unit, called the *henry,* is defined by reference to the volt and ampere: it is *the inductance of a coil in which a change of one ampere per second produces an e.m.f. of one volt.* Since the volt is 10^8 e.m.u. and the ampere is 10^{-1} e.m.u. the henry is 10^9 e.m.u.

If a varying current I_1 in one coil (the primary) induces an *e.m.f.* E_2 in another coil (the secondary)

$$E_2 = -M\frac{dI_1}{dt}$$

The constant M, called the *mutual inductance* of the coils, is the same whichever coil is the primary. It is also measured in henries.

494. Effect of a Large Self-inductance.—*A* is a 110-volt lamp (Fig. 396) in series with a resistance *R* and in parallel with a coil of large self-inductance *L* and small resistance, for example an electromagnet. If a 110-volt source of e.m.f. is connected in through the key *K*, the lamp flashes up for a moment and then becomes dim. When the circuit is broken the lamp again flashes up for a moment before going out. If *L* is replaced by a carbon rheostat, (that is, a resistance having no self-inductance) adjusted so that the steady light from the lamp is about the same as before, no flashes at make or break are observed.

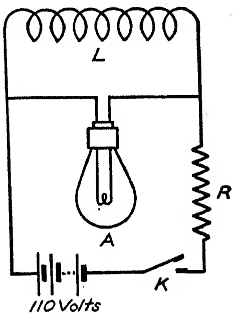


FIG. 396.

These effects are explained by considering the self-induced e.m.f.'s in *L*. At make of the current the self-induced e.m.f. in *L* prevents the current building up rapidly in *L* and it goes, practically all, into *A*. But after a second or two the currents become steady, there is then no induced e.m.f. and *L* carries a part of the current that depends only on its resistance. When the current is broken by opening *K*, the rapid fall of the current in *L* produces a large induced e.m.f. in the circuit formed by *L* and *A* in series, and a large momentary current flows through *A*.

495. Rise of Current in an Inductance.—Let us now examine in some detail the process of growth of a current in a circuit in which there is a coil of self-inductance *L* (Fig. 397). The current is started by closing a key and so throwing into the circuit an e.m.f. *E*, which we call the applied e.m.f. But as the current starts to grow an induced e.m.f., opposite in direction to *E*, is created in *L*. At a time *t*, when the current at the moment is *I*, the induced e.m.f. is $-L\frac{dI}{dt}$. Then the effective e.m.f. at the moment is $E - L\frac{dI}{dt}$. By Ohm's law the e.m.f. in a circuit equals the product of current by resistance. Hence

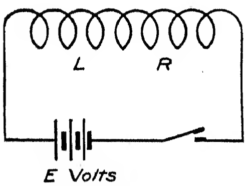


FIG. 397.

$$E - L\frac{dI}{dt} = RI$$

From this it can be deduced that the relation between *I* and *t* is

$$I = \frac{E}{R}\left(1 - e^{-\frac{R}{L}t}\right)$$

If the reader is not familiar with the method of deriving this equation, he may be able to show that the second equation gives the first by differentiation.

If we now plot I as a function of t we get curves such as are shown in Fig. 398. Curve A represents the case in which L is small, the current growing rapidly. Curve B is for a case in which L is large and the growth is slow.

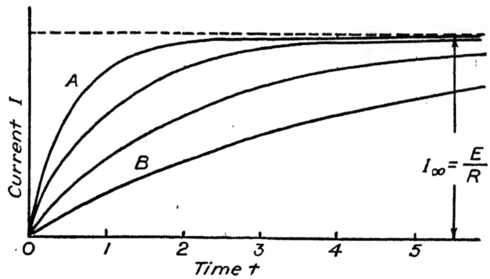


FIG. 398.

496. Energy of a Current.—Establishing a current requires work, for the opposition of self-induction, which is of the nature of inertia, has to be overcome. To find an expression for the work done in establishing a current I, let us denote the current at a time t, while it is still growing, by i. In an element of time dt the quantity that flows is idt. The magnitude of the opposing e.m.f. of self-induction at time t is $L\frac{di}{dt}$. The product of these, which, by the definition of e.m.f. (§422) is the work done, is $Lidi$. To get the whole work we must sum these elements (integrate) up to the full current I. The result is $\frac{1}{2}LI^2$. It is interesting to note the similarity of this expression and the formula $\frac{1}{2}mv^2$ for the kinetic energy of a moving body.

497. Inductance of a Solenoid.—If there is a current of I e.m.u. in a solenoid of n turns per cm. it produces within the solenoid a magnetic field of strength H where (§431)

$$H = 4\pi nI \text{ oersteds, } (I \text{ being in e.m.u.})$$

If the permeability of the medium inside the solenoid is μ, the magnetic flux in the medium is (§440)

$$\phi = \mu HA = \mu 4\pi nIA \text{ maxwells}$$

where A is the area of cross-section of the solenoid. Denote the length of the solenoid by l. Then the total number of turns N equals nl and the self-inductance of the coil is (§493)

$$L = \frac{N\phi}{I} = \mu 4\pi n^2 Al \text{ e.m.u.}$$

In this reasoning we have assumed that the flux from each turn links with every other turn. As this is not strictly true, a correction for the ends of the coil (which we need not consider here) may have to be applied. Moreover if the medium is iron μ is not constant and the expression for L applies only to a particular value of μ.

498. Energy in a Magnetic Field.—Consider again the case of the solenoid of §497. Work of the amount $\frac{1}{2}LI^2$ has to be done in starting a current I in it and there is therefore this amount of energy in the magnetic field created. From the results of §497 we readily get

$$W = \frac{\mu H^2}{8\pi}Al$$

Now Al is the volume of the magnetic medium. Hence the magnetic energy per unit volume in it is $\mu H^2/8\pi$. This corresponds exactly to the expression $KE^2/8\pi$ for the electrical energy per unit volume in an electric field (§397).

499. Induction Coil.—The induction coil is an apparatus for producing pulsating currents or discharges of high e.m.f. in a second-

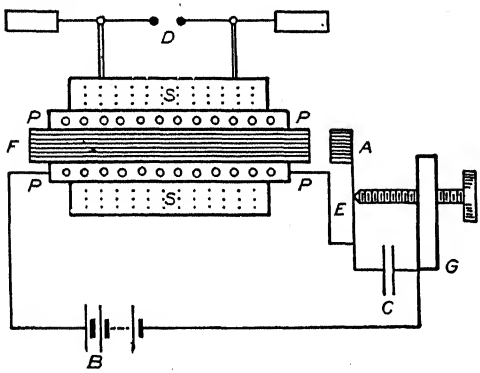


FIG. 399.

ary circuit S by making and breaking a current in a primary circuit P. P consists of a few turns of heavy wire wound on a laminated soft iron core F. S consists of a large number of turns of fine wire. The making and breaking of the current is automatic; the current in P magnetizes the core F and F then attracts a piece of soft iron A and so breaks the current at E; but when the current is broken and F is no longer magnetized, A springs back and closes the primary circuit again, and so on. The contacts at E are of platinum to withstand wearing away; and, to prevent heavy sparking and make the break more sudden and the make more gradual, a condenser is

placed in parallel with the spark gap. For large induction coils the mechanical interrupter shown in the figure is replaced by an electrolytic interruptor (Wehnelt) or a rotating jet of mercury.

Fig. 400 shows a typical case of the changes of primary current (solid curve) and secondary current (dotted curve) in a cycle. The primary current builds up slowly, owing to the self-inductance of its circuit, and the secondary current is small; but the break of the primary is sudden and the secondary e.m.f. is therefore large.

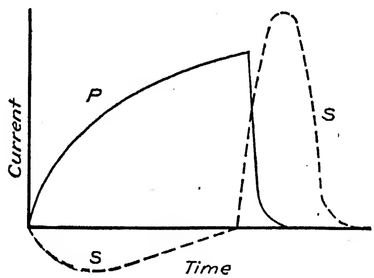


FIG. 400.

500. Earth Inductor.—Induction of a current in a coil rotating in the earth's magnetic field is made use of in the earth inductor (Fig. 401). Denote the area of the coil by A and the horizontal component of the earth's field by H. Let us suppose that the axis of rotation is vertical and that the plane of the coil is perpendicular to the magnetic meridian. There is then a magnetic flux HA through the coil, the vertical component V of the field producing no flux. Rotating the coil through 90° withdraws this flux and rotation through another 90° puts it in again but in the opposite direction through the coil, so that the whole change of flux is $2HA$, and the induced quantity of electricity, which can be measured by a ballistic galvanometer in the circuit, is given by (§487)

$$Q = \frac{2HA}{R}$$

where R is the resistance of the coil and galvanometer.

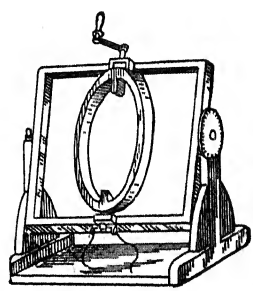


FIG. 401.

Next suppose the plane of the coil to be horizontal. The flux through it is then due to the vertical component, V, of the earth's magnetic field. When the coil is rotated through 180° the total change of flux through it is $2VA$ and, if the induced quantity is Q',

$$Q' = \frac{2VA}{R}$$

Hence

$$\frac{Q'}{Q} = \frac{V}{H} = \tan \phi$$

where ϕ is the angle of dip of the earth's magnetic field (§414).

501. Induced Currents in Testing Iron.—For finding the relation between magnetizing field H and induction B in iron (§444) a method founded on Faraday's experiment with a primary coil and a secondary coil on an iron ring has been found the most satisfactory. First a primary coil of n turns per cm is wound on the whole ring (Fig. 403). Over this is wound a secondary coil Sc of N turns. A ballistic galvanometer is put in series with the secondary, and its throws on change of current in the primary are observed. The constant of the galvanometer is found by means of an earth inductor I in the secondary circuit.

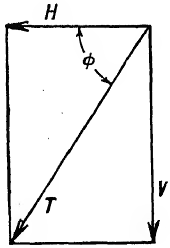


FIG. 402.

As an example of the use of the method we shall consider how it might be applied to find the hysteresis loop of §444. H is increased by steps and the corresponding increases of B are found. Now the value of H for any current i is given by the $4\pi nI$ law (§431). Hence when i is increased from i_1 to i_2 so that H increases from H_1 to H_2

$$H_2 - H_1 = 4\pi n(i_2 - i_1)$$

To find the increases of B, we go back to the relation $Q = (\phi_2 - \phi_1)/R$ (§487). In this $(\phi_2 - \phi_1) = N(B_2 - B_1)A$ where A is the area of cross section of the

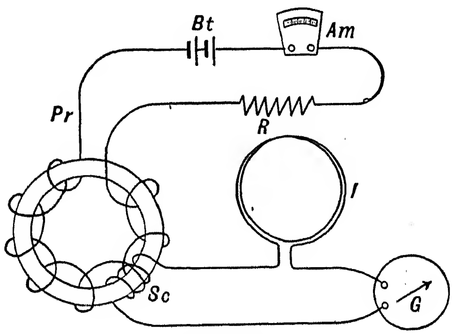


FIG. 403.

ring. Hence

$$Q = \frac{N(B_2 - B_1)A}{R}$$

Now Q can be found from the throws of the ballistic galvanometer, and then this equation gives the increase of induction $B_2 - B_1$ produced by the increase

of magnetizing field $H_2 - H_1$. These operations, carried out step by step, give the whole loop.

502. Telephone Transmitter and Receiver.—In a telephone sound waves, impinging on a thin disk, produce vibrations that affect the electric current in the line; and, at the receiving end, these variations of current act on another disk so as to reproduce sound waves.

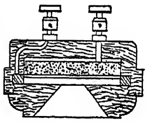


FIG. 404.

The ordinary transmitter (Fig. 404) makes use of the changes of resistance of granular carbon, packed between two small carbon plates, when one of the plates vibrates. These variations of resistance in the transmission line, which contains a battery or generator, cause corresponding variations in the current transmitted. The variations of a carbon contact that are made use of here are well illustrated by the action of a carbon rod resting on two other carbon rods (Fig. 405), all three being in series with a battery and a telephone receiver. The slightest jar of the support causes a sound in the receiver (a galvanometer can also be used). In another form of transmitter

COMPARISON OF ELECTRICAL AND MAGNETIC SYSTEMS OF UNITS

	Electrostatic	Electromagnetic	Practical
Charge	Defined from Coulomb's Law (§369) 1 e.s.u.	Defined from $Q = It$ 1 e.m.u. $= 3 \times 10^{10}$ e.s.u.	Defined from $Q = It$ 1 coulomb $= \frac{1}{10}$ e.m.u. $= 3 \times 10^9$ e.s.u.
Current	(Not used)	Defined from Amperes law (§427) 1 e.m.u.	Defined from e.m.u. of current 1 ampere $= \frac{1}{10}$ e.m.u.
Potential difference	Defined from $W =$ P.D. $\times Q$ 1 e.s.u.	Defined from $W =$ P.D. $\times Q$ 1 e.m.u. $= 1/(3 \times 10^{10})$ e.s.u.	Defined from e.m.u. of P.D. 1 volt $= 10^8$ e.m.u. $= \frac{1}{300}$ e.s.u.
Capacity	Defined from $Q = CV$ 1 e.s.u.	Defined from $Q = CV$ 1 e.m.u. $= 9 \times 10^{20}$ e.s.u.	Defined from $Q = CV$ 1 farad $= 10^{-9}$ e.m.u. $= 9 \times 10^{11}$ e.s.u.
Inductance	(Not used)	Defined from $E = -LdI/dt$ 1 e.m.u.	Defined from $E = -LdI/dt$ 1 henry $= 10^9$ e.m.u.

the diaphragm on which the sound falls carries a small coil that is in a magnetic field. Currents are induced in the coil and act on the line through a transformer. Still another transmitter uses the varying capacity of a small condenser, one plate of which is attached to the diaphragm.

The ordinary receiver (Fig. 406) has a permanent magnet of a U-shape, around the poles of which a coil in series with the transmission line is wound, the

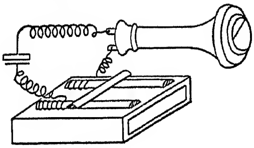


Fig. 405.

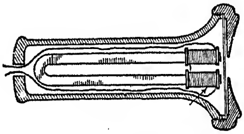


Fig. 406.

windings on the two poles being in opposite directions. Thus to the field of the magnet is added a field that varies as the current varies. A soft iron disk close to the magnet, but not touching it, always tends to move into the strongest part of the field between the poles. Being restrained by its own resistance to deformation, it vibrates, keeping step, fairly closely, with the varying current, and so reproducing the original sound. In calling, the operator uses an alternating current that acts through a condenser which is in parallel with the transmitter.

GENERATORS AND MOTORS

503. Alternating Current Generators.—In the production of large currents for industrial purposes use is made of the induction of an e.m.f. in a coil, called an *armature*, that rotates between the poles of an *electromagnet*. The armature is kept in motion by a steam engine or turbine, and the result is the transformation of mechanical energy into the energy of an electric current.

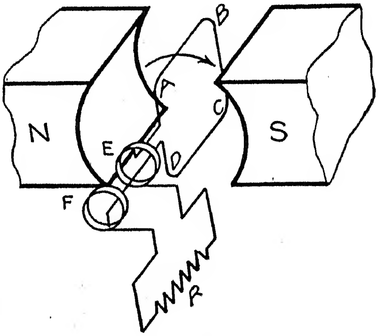


Fig. 407.

Let us begin with the alternating or A.C. generator, inasmuch as e.m.f.'s reversing in direction are naturally produced in a rotating coil. The ends of the armature *ABCD* are connected to *slip-rings* *E* and *F* (Fig. 407) that rotate with the armature, and connection with the external circuit is through carbon *brushes* that are pressed against the slip-rings. As the armature rotates the magnetic flux through it at any moment is proportional to the projection of its area on a plane perpendicular to the lines of force of the field; and

the e.m.f. induced at any moment (§486) is the rate of change of this flux. The flux is a maximum when the armature is perpendicular to the field and the rate of change of flux is then zero. On the other hand, the flux is zero when the armature is parallel to the field and the rate of change of flux is then a maximum. From these relations we see, at least in a general way, that a curve of induced e.m.f. plotted against flux, must be a simple harmonic curve (Fig. 147), and this will be considered more closely later.

The frequencies of alternating currents are standardized in this country at 60 cycles and 25 cycles per second, the former being used for lighting to avoid perceptible flicker. To get sufficient e.m.f. without a high rate of rotation, the *multipolar generator* (Fig. 408*a*)

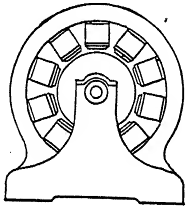


Fig. 408*a*.

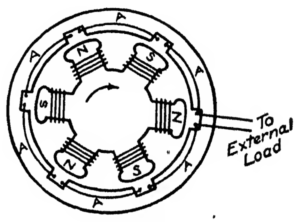


Fig. 408*b*.

was invented. It is equivalent to several simple generators in series and kept in step. The armature consists of a number of coils on an iron core and connected in series, and there is an equal number of field magnets. The e.m.f.'s produced by the coils reach their maxima at the same time, and the resultant e.m.f. at any moment is the numerical sum of the separate e.m.f.'s. When a number of e.m.f.'s are related in this way, they are described as being *in phase*, that is, always in the same part of a cycle. The simple a.c. generator and the multipolar generator are called *single phase* generators to distinguish them from some other generators that we shall consider later.

While we started by speaking of the magnets as at rest and the armature in rotation, in reality it is found better to keep the armatures, which are relatively more complex, at rest and the field magnets in rotation (Fig. 408*b*). In any case the stationary part is called the *stator* and the rotating part the *rotor*.

Denote the angular velocity by ω and the period by τ, so that $\omega = 2\pi/\tau$. If the lines of force of the field (Fig. 407) are horizontal the flux through the armature when it is vertical is μHA (§440). If

time is reckoned from this position, at time t the armature makes an angle ωt with the first position and the flux through it is $\mu HA \cos \omega t$. The rate of increase of this expression (as found by differentiating) is $-\mu HA\omega \sin \omega t$. Hence the e.m.f. at time t or the rate of decrease of the flux (§486) is

$$E = E_m \sin \omega t$$

where E_m is the maximum value of the e.m.f. and is equal to $\mu HA\omega$. This is the equation of the continuous curve in Fig. 410. The result applies also to a multipolar generator, but E_m becomes $n\mu HA$, where n is the number of coils in the armature.

504. Alternating Current Circuit.—A single phase alternator in action in a circuit containing any number of resistances and self-inductances (Fig. 409) must produce an alternating current; and the frequency of the current or the number of cycles per second must be the same as the frequency of the alternating e.m.f. Curves representing e.m.f. and current will be similar in form (Fig. 410). But there must be a difference of phase between the e.m.f. and the current. To see the reason for this, think again of what happens at starting or stopping of a direct current from a battery. As the current begins to rise, self-induction sets in, producing an opposing e.m.f. and so delaying the growth of the current; and when the current is stopping self-induction tends to prolong it. In both cases self-induction causes a *lag* of current behind applied e.m.f. Exactly the same is true of each half cycle of an alternating current, for it also is a rise and a fall of current. The whole effect is a uniform lag of the current behind the applied e.m.f.

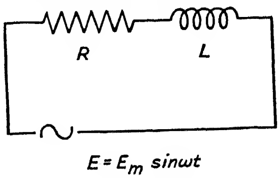


FIG. 409.

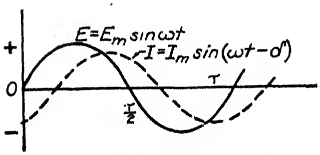


FIG. 410.

To express these relations between e.m.f. and current by formulas, let us denote the lag of current behind e.m.f. by an angle δ subtracted from ωt. Then, as an expression for the current I, we write

$$I = I_m \sin (\omega t - \delta)$$

where I_m is the maximum value of the current, and δ is supposed to be (like ωt) in radians. For example, consider two special cases:

(1) No circuit is entirely free from self-induction, but if there are no electromagnets or solenoids present the self-inductance will be very small. In this case the lag due to self-induction is very slight and δ is practically zero (Fig. 411).

(2) If, on the other hand, the resistance is very small and the self-inductance of the circuit large the lag of the current behind the e.m.f. approaches $\pi/2$ or $\tau/4$, a quarter of a cycle, in time (Fig. 412).

505. Application of Ohm's Law.—Let us now consider how Ohm's law applies to e.m.f. and current in an alternating current circuit. Since the applied e.m.f., $E_m \sin \omega t$, is opposed by the e.m.f. of self-induction, LdI/dt, the effective or resultant e.m.f. at any moment is $(E_m \sin \omega t - LdI/dt)$. By Ohm's law, this e.m.f. divided by the resistance R of the circuit is equal to the current I. We thus get, by rearranging terms,

$$L\frac{dI}{dt} + RI = E_m \sin \omega t$$

We shall see presently how this fundamental equation can be put into a form more suitable for calculations. In the meantime let us consider again the two special cases referred to above.

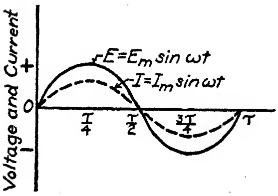


FIG. 411.

(1) If the self-inductance of the circuit is negligibly small, we may put $L = 0$. Then

$$I = \frac{E_m}{R}\sin \omega t = \frac{E}{R}$$

Hence for an inductionless circuit there is no difference of phase between current and e.m.f. (Fig. 411).

(2) If the resistance of the circuit is negligibly small, we may put $R = 0$. We then have

$$\frac{dI}{dt} = \frac{E_m}{L}\sin \omega t = \frac{E}{L}$$

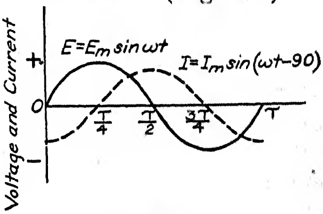


FIG. 412.

Now when I is at a maximum, dI/dt is zero and the equation shows that E is then also zero. Thus the current is 90° or a quarter of a cycle behind the e.m.f. in phase (Fig. 412).

506. Impedance of an A.C. Circuit.—The equation that expresses Ohm's law for a circuit containing resistances and inductances (§505) is not of such a form that it can be used for calculating I, since it also contains the rate of increase of I, that is, dI/dt. It is, however, not difficult to derive from it an expression for I that is more directly useful, though we shall have to omit the purely mathematical steps and merely show how the result is used.

When an alternating e.m.f. $E_m \sin \omega t$ is applied to a circuit of resistance R and self-inductance L (Fig. 409) the current I at any moment is expressed by

$$I = \frac{E_m \sin (\omega t - \delta)}{Z}$$

where Z is called the *impedance* of the circuit and δ, as already explained, is the lag of current behind applied e.m.f. Now the mathematical deduction, which we have omitted, shows that

$$Z = \sqrt{R^2 + L^2\omega^2} \qquad \text{and} \qquad \tan \delta = \frac{L\omega}{R}$$

Evidently what we have called impedance, Z, plays about the same part in an alternating current circuit as resistance does in a direct

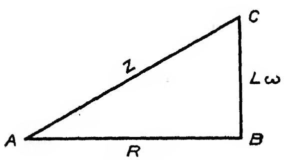


FIG. 413*a*.

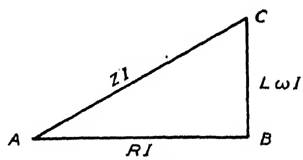


FIG. 413*b*.

current circuit. $L\omega$ is called the *reactance* of the circuit or, more exactly, the induction-reactance to distinguish it from another reactance when there are condensers in a circuit (§508).

Now it will be noticed that Z is related to R and $L\omega$ exactly as the hypotenuse to the sides of a right-angled triangle (Fig. 413*a*), and this is a very useful and suggestive relation. From this another useful relation follows. If we multiply R, $L\omega$, and Z by I, RI is the potential drop in the circuit due to the resistance, $L\omega I$ that due to the inductance and ZI the total drop, and these three are also related to one another as the sides of the triangle. Since $\tan CAB =$

$BC/AB = L\omega/R$, CAB is the angle of lag δ. It will be somewhat clearer why a right-angled triangle is so exactly suitable for expressing these relations if we note again that (Fig. 412) an alternating current in a circuit with inductance, but practically no resistance, lags 90° behind the impressed e.m.f., so that we would expect $L\omega$ to combine with R in some such right-angle construction.

507. Average Value of an Alternating Current.—In the preceding I has always meant the current at a certain moment, but it is a varying current, and for many purposes what is needed is its average value. Now "average" is an indefinite term until we specify how the average meant is to be calculated, and this depends on the purpose in averaging.

Alternating currents are used for transmitting energy. The energy delivered by the current may take many forms; but, as all of these are, by the conservation of energy, interchangeable and equivalent, it will be sufficient to consider one form only, say heat. Suppose, then, that an alternating current I flows through an inductionless resistance R for t seconds, where t is a whole number of cycles. It will simplify expressions, without affecting the result, if we take R as 1 ohm and t as 1 second. Now a direct current, say I_e, flowing through the same resistance for the same time would produce I_e^2 joules (§423). In the case of the alternating current let us divide the one second of flow into a very large number n of very small equal intervals, each $1/n$ second. When the strength of the alternating current is I, it produces in one such interval I^2/n joules. To get the whole heat we must sum I^2/n for all the values of I. This evidently gives the time-average of I^2. Hence if I_e is the direct current equivalent to the alternating current, I_e^2 = time-average of I^2, or $I_e = \sqrt{(I^2)_{av}}$. I_e is called the root-mean-square value of I.

To find I_e for an alternating current, $I_m \sin(\omega t - \delta)$, let us, for brevity, write α for $(\omega t - \delta)$. Now $\sin^2 \alpha + \cos^2 \alpha = 1$. The average values of $\sin^2 \alpha$ and $\cos^2 \alpha$ in a cycle (360°) are evidently equal and each must therefore be $\frac{1}{2}$. Hence $I_e = I_m/\sqrt{2} = 0.707 I_m$. Similarly, the effective value of an alternating e.m.f. is $0.707 E_m$. A.C. ammeters and voltmeters are graduated in effective values. Thus a 110-volt alternating e.m.f. has a maximum value of 155 volts, and would produce the same amount of heat per second in an inductionless resistance as would a 110-volt direct current applied to the same resistance.

508. Condenser in an A.C. Circuit.—A direct current cannot continue to flow through a condenser, but an alternating current can. The plates become charged alternately positive and negative; and, when charged, they discharge back into the circuit. The condenser acts somewhat like a spring between a steam engine and a load, whereas an inductance is analogous to a fly-wheel. When the condenser is charged, it exerts an e.m.f. in the circuit, just as any two oppositely charged bodies produce an electric field. We could now write down and solve an equation for the whole process, just as in the case of a circuit with resistances and inductances. We would then find that, in calculating the impedance Z, we must take account of a *capacity reactance.* While the derivation must be omitted, there is one important result that can readily be understood.

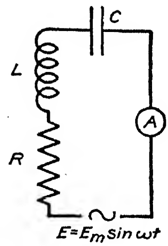


FIG. 414.

For simplicity, suppose the resistance and the inductance of a circuit containing a condenser are negligibly small. It is found, by the method we have sketched, that in this case the current is 90° out of phase with the applied e.m.f.; but it is now the current that is ahead in phase, so that δ is a *lead*, not a lag (Fig. 415). To see why this must be so, note that, when the applied e.m.f. begins to fall from its positive maximum, the condenser, now fully charged, begins to discharge in the *opposite* direction.

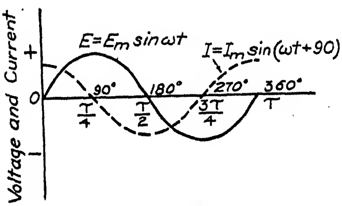


FIG. 415.

509. Power in an A.C. Circuit.—In a direct current circuit, power, or the rate at which a current can do work of some kind, is equal to EI, where E and I are steady values of e.m.f. and current (§422). The same expression holds true at any instant in an alternating current circuit. But this instantaneous value of EI cannot be used directly for calculating the *average power* P in an alternating current circuit, for E and I are varying rapidly. P evidently depends on the phase relation of E and I. If they are 90° apart in phase (Fig. 412) each is zero when the other is a maximum and in this case P is zero; when E and I agree in phase (Fig. 411) P is a maximum. A general expression for P can be derived mathematically from the instantaneous EI, namely

$$P = \tfrac{1}{2}E_m I_m \cos \delta$$

where E_m and I_m are maximum values and δ is the lag or phase difference of E and I; and, if we substitute effective values of E and I (§507)

$$P = E_e I_e \cos \delta$$

This result might also be derived from the graphical construction or vector diagram (Fig. 413); for $E \cos \delta$ is the component of E which is in phase with the current I, and it is the only component that is effective as regards power.

The preceding has been on the assumption that the applied e.m.f. is of the simple form, $E_m \sin \omega t$, sometimes called a sinusoidal or simple harmonic e.m.f.

In many cases the e.m.f. is more complex, consisting of terms like $E_m \sin \omega t$ but with different values of ω (like harmonics from a piano wire §575); and the graph of e.m.f. is then not sinusoidal. In this case ω and $\cos \delta$ (which depends on ω) have no definite meanings; and, in the expression for P, $\cos \delta$ is then replaced by $\cos \theta$, which is called the *power factor* and has to be determined experimentally.

Power in an a.c. circuit is measured by a wattmeter of the same general design as that used for d.c. (§438) but graduated so as to measure average power in an a.c. circuit.

We have been assuming in this discussion that the resistance R of an a.c. circuit is the same as its d.c. resistance. If there is any iron in the circuit then there are hysteresis and eddy current losses in the iron. Energy is used up in such cases which goes into heat. The effective resistance R for an a.c. circuit must include such losses. Hence its value in ohms is given by

$$R = \frac{P}{I_e^2}$$

where P is the power in watts expended in the entire circuit in which there is an alternating current of effective value I_e.

510. Transformers.—One reason why alternating currents have largely replaced direct currents for practical use is the ease with which an alternating voltage can be raised or lowered as desired. Since power depends on the product of voltage and current, it is more economical to transmit it at a high voltage with a small current, for less power (I^2R) is wasted in the lines. Thus the voltage needs to be raised for transmission and lowered again, as may be necessary, for use at the other end of the line. Both of these operations require *transformers*.

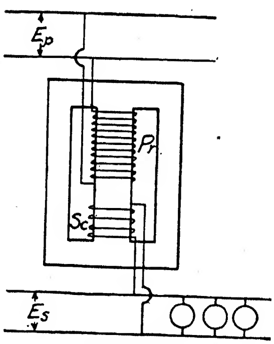


FIG. 416.

An alternating current transformer is essentially Faraday's iron ring with a primary and a secondary winding (§484). An alternating current in the primary sets up an alternating magnetic flux in the ring, and this produces an alternating e.m.f. in the secondary. In a solid iron ring hysteresis and eddy currents would produce heat and loss of useful energy. To reduce this, the ring is made of laminated iron and constructed so as to reduce magnetic leakage to a minimum. The result is that the efficiency of most transformers is quite high, ranging (for full load) from 95 to 99 per cent.

In an ideal transformer there would be exactly the same alternating magnetic flux through primary and secondary. If there were the same number of turns in the two windings, the effective e.m.f., say E_2, in the secondary would be the same as the effective e.m.f., E_1, in the primary. With N_1 turns in the primary and N_2 in the secondary, the e.m.f. in the secondary would be $(N_2/N_1)E_1$ or we would have

$$\frac{E_2}{E_1} = \frac{N_2}{N_1}$$

In real transformers this relation is nearly correct when there is no load on the secondary; but when there is a load the relation is only roughly approximate.

For an ideal transformer, of 100 per cent. efficiency as regards power, $E_1I_1 \cos \delta_1$ for the primary would be equal to $E_2I_2 \cos \delta_2$ for the secondary. Also, by a self-regulating action-and-reaction through the iron core, δ_1 and δ_2 would be equal. We would then have

$$\frac{I_2}{I_1} = \frac{E_1}{E_2}$$

In actual transformers this relation also is only a rough approximation.

The primary of a transformer is usually permanently connected to a source of a.c. while the secondary is closed through a switch only when power is needed from the transformer. When the switch in the secondary is open, so that there is no current in the secondary, there is a small magnetizing current present in the primary. This magnetizing current is practically 90° out of phase with the applied voltage and no power is wasted in the primary. All the flux produced by the magnetizing current cuts the primary coil and sets up a back e.m.f. (LdI/dt) which balances the applied voltage at every instant.

When the switch in the secondary circuit is closed and there is a current in the secondary, this current produces magnetic flux opposite in direction to that produced by the current in the primary (Lenz's Law). Thus the magnetic flux and back e.m.f. in the primary coil is reduced so that now there is a larger current in the primary circuit. Hence the current in the primary increases when the current in the secondary increases.

511. Two Phase Alternator.—We shall see later that an important use can be made of two alternating e.m.f.'s in two separate lines, provided the e.m.f.'s are 90° apart in phase, so that one is at its maximum when the other is zero (Fig. 417). If the equation for one of these e.m.f.'s is

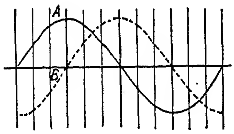


FIG. 417.

$$E = E_m \sin \omega t$$

the equation for the other may be written in the form

$$E' = E_m \cos \omega t$$

since $\sin(\omega t + 90°) = \cos \omega t$. A very simple way of providing two such e.m.f.'s is to have an armature consisting of two entirely separate coils on the same frame and rotating between the poles of a single electromagnet (Fig. 418). This is the simplest form of a *two phase alternator*. As each coil has its own two terminals, it might be expected that four transmission lines would be required for use of both e.m.f.'s at a distance, but it is found that one line can be shared by the two phases so that only three lines are necessary. Each coil with the electromagnet constitutes a single phase generator and can be used as such. As in the case of the single phase alternator, the multipolar principle is used to get higher e.m.f.'s.

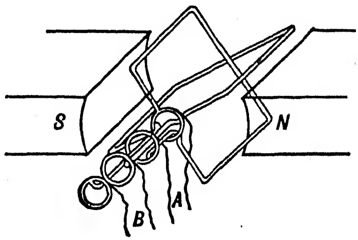


FIG. 418.

512. Three Phase Alternators.—Three separate coils on the same frame, each two 120° apart, and rotating between the poles of a single magnet constitute, in its simplest form, a *three phase alternator*. The e.m.f. curves of the three phases are as represented in Fig. 419. In actual construction the multipolar principle is made use of, and the armature with its three coils is the stator, the field being the rotor. At first sight this might be expected to involve an intolerable number of terminals and line wires, or six lines in all. As a matter of fact the number can be reduced to four and this *three-phase, four-wire* system is now the usual way of distributing light and power in cities. The number

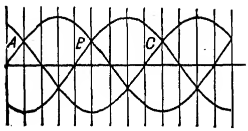


FIG. 419.

of lines can for many purposes be reduced to three, but we cannot explain here the ingenious methods for connecting coils or lines together to produce this result. The advantages of three-phase transmission of power as compared with single-phase transmission depend on the superiority of the motor which it uses (§516) and a great saving of copper for the same power transmitted. For these reasons it has practically replaced other methods for transmitting power.

513. Direct Current Generator.—The direct current generator, giving a current that is always in the same direction and does not reverse, was the earliest form of generator. It is still needed for many special purposes, such as charging storage batteries, electroplating, electrolysis in its various applications, as well as for running street cars and elevators, where the superior starting power of the direct current motor is important.

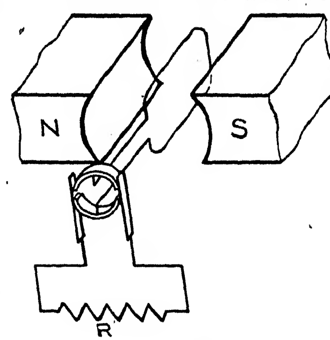


FIG. 420.

In principle a direct current generator (Fig. 420) is an alternating current generator with an automatic device for reversing connections to the transmission line for every alternate half cycle, so that the e.m.f. curve of Fig. 410 becomes the "rectified" curve of Fig. 421. The rectifying device consists in replacing the two slip rings of Fig. 407 by a single slit ring, the two halves being joined to the ends of the armature coil. This is called the *commutator* of the machine. The brushes are adjusted in

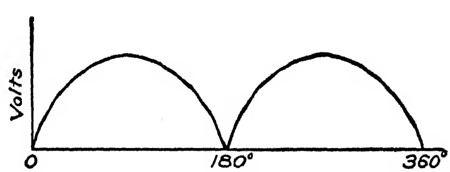


FIG. 421.

position so that the connections to the external circuit are reversed just when the current in the coil would itself reverse.

The pulsating e.m.f. from a single coil, varying from zero to a maximum, could not be used commercially. If, however, the multipolar principle were applied, the single coil being replaced by a

number of coils with a corresponding number of magnets and each coil having its own pair of segments in the commutator ring, the result would be as indicated in Fig. 422. In practice the separate coils are replaced by segments of a continuous winding of insulated wire on an iron core, the wire being tapped at regular intervals for connection to the commutator segments. If the iron core is a ring (Fig. 423) the winding of the armature is called "ring-winding"; if it is a hollow cylinder the winding is called "drum-winding."

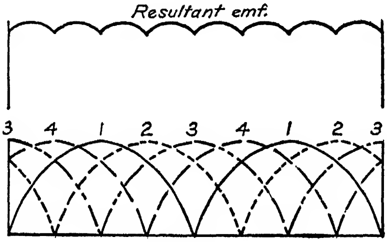


FIG 422.

Direct current generators are also classified according to how the current to excite the field magnets is derived from the generator itself. If the coils of the field magnet are in series with the main circuit, the generator is called a "series generator"; if they are in parallel with the main circuit, the generator is a "shunt generator." These are, however, all very general ideas; the details of structure of such generators are too complex for consideration here.

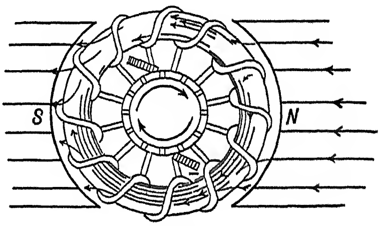


FIG. 423.

514. Electric Motors.—An electric generator transforms mechanical energy into the energy of an electric current and an electric motor reverses the process, transforming the energy of the electric current into mechanical energy. In both the most elementary part is a conductor at right angles to a magnetic field. In the generator the conductor is moved across the field and a current is induced in it; in the motor a current in the conductor makes it move across the field; the direction of motion is opposite in the two cases if the direction of the current is the same. From this relation it would appear that a generator might also serve as a motor, running backward, when fed with a current from a similar machine as generator. In a general way this is true, but in details the two machines may differ considerably, according to their purposes.

A direct current motor is very much like a direct current generator, except for minor details. It starts of itself when direct current

is applied to it and it can run at different speeds; but some additional resistance must be thrown in at the start to prevent an excessive current in the armature before the back e.m.f. of self-induction is built up. Also a single phase alternator can serve as a motor, but only when it has been brought up to the same speed as the alternator that generates the current, so that their rates of alternation of current agree. Such a motor is called a *synchronous motor*. To bring it up to the synchronizing speed before it is connected to the line, some auxiliary motor or winding is necessary. A small motor, called a universal motor, operates on both direct currents and alternating currents and is used for sewing-machines, electric fans, vacuum cleaners, etc. It is essentially a series wound direct current motor. When an alternating current is applied to it the current is reversed simultaneously in armature and field magnets, so that the direction of the torque is always the same.

In all motors the mechanical work which the the motor can do is equal to the work done against the back induced e.m.f., that is, the electrical energy expended against this e.m.f.

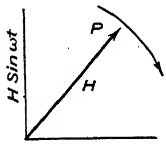


FIG. 424.

515. A Revolving Magnetic Field.—A magnetic field has a certain magnitude called the strength H of the field and it is in a certain direction. It is therefore a vector quantity and can be represented by a directed line OP in a diagram (Fig. 424). Suppose now that H is constant but that OP revolves steadily in a plane with an angular velocity ω. At any moment OP may be resolved into a component $H \cos \omega t$ along OX and a component $H \sin \omega t$ along OY. Conversely two component fields at right angles and varying as $H \cos \omega t$ and $H \sin \omega t$ will give a revolving magnetic field of strength H. Now this relation between two magnetic fields is the same as the relation between the two currents that can be produced by a two phase alternator (§511). Hence, if we use these two currents in two electromagnets at right angles (Fig. 425), the resultant magnetic field will be a revolving field.

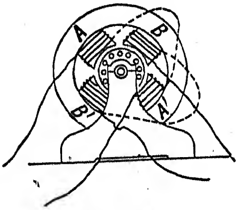


FIG. 425.

A revolving magnetic field can also be obtained by using the three currents from a three phase alternator in three electromagnets inclined at 120° to each other. This result can be worked out by resolving the components along two axes at right angles; but it seems so nearly obvious from the symmetry of the arrange-

ment and the way in which the magnets come to their maxima in succession that we shall assume it here.

516. The Induction Motor.—The importance of the revolving magnetic field is due to its use in a remarkable motor called the *induction motor*

The principle at the basis of its action is illustrated by an experiment made by Arago more than a hundred years ago but not explained satisfactorily until Faraday discovered the principle of the induction of currents. If a conducting disk is pivoted above a magnet that is kept in rotation as in Fig. 426, the disk also rotates, though at a lower speed than the magnet. If the magnet is free to rotate and is initially at rest and the disk is driven, the magnet also rotates, but at a lower speed than the disk. The explanation is that currents are induced in closed circuits in the disk, as its parts cut across the lines of force of the magnet, and these currents are in such directions as to oppose relative motion of disk and magnet (Lenz's law). If there are radial slots in the disk, there is no dragging action. In the induction motor the armature, which is also the rotor, consists essentially of a number of parallel copper rods fastened to two copper end plates (Fig. 427). Because of this peculiar form it is often called a "squirrel cage" armature. When it is in a revolving magnetic field (Fig. 424) the magnetic lines of force of the revolving field cut across the rods and currents are induced in the rods. The currents are, by Lenz's law, in such a direction as to oppose by their magnetic fields the *relative motion* of field and armature; and, as the rotation of the field is kept up, the armature starts to follow the field in rotation. The motor is therefore self-starting.

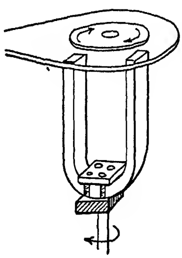


FIG 426

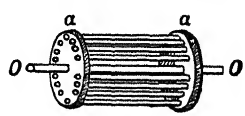


FIG. 427.

When it has started the armature does not catch up with the field; there is a "slip" or difference of speed between them; for, if it did revolve as fast as the field, there would be no relative motion between them and therefore no induced currents to keep the armature running. If it is given an impulse to bring it up to synchronism with the field, it drops back at once, until the slip between them is just sufficient to produce the currents needed for the load on the armature. Thus the motor is self-adjusting as to speed and

can, within certain limits, be used at any speed and for a widely varying load.

When the motor is running and one of the line wires is disconnected, it will continue to run, but it will not then start without an impulse. It can be reversed by reversing connections to the field magnets, so as to reverse the direction of revolution of the field. The fact that the induction motor has no slip rings or commutators helps to make it more rugged and durable than motors of other types.

ELECTROMAGNETIC WAVES

517. Displacement Currents.—When an alternating e.m.f. is applied to a condenser, alternations of charges on the plates are produced. At a moment when the charges are $+Q$ and $-Q$, the electric field strength in the dielectric in the condenser is (§391)

$$R = \frac{4\pi}{KA}Q$$

where A is the area of each plate and K is the dielectric constant of the medium. The current I that is flowing into the condenser is the rate at which Q is increasing, which we denote by dQ/dt. Hence

$$I = \frac{dQ}{dt} = \frac{KA}{4\pi}\frac{dR}{dt}$$

where dR/dt means the rate at which R is increasing. Now this current I is the current in the transmission line. It carries energy with it, and this energy travels through the dielectric in the condenser, and I is proportional to dR/dt. We find it necessary, therefore, to think of a varying electric field strength R in a dielectric as being a form of electric current. It is called a *displacement current* in the dielectric.

We can account, in part, for a displacement current in a dielectric by thinking of the motions of electrons that are *bound* to atoms, not free electrons as in conduction currents. These bound electrons remain in the atoms to which they belong, revolving in some way around the nuclei of the atoms; but their orbits must be distorted when they are in an electric field, so that they are all *displaced* somewhat in the direction of the field. When the field is an alternat-

ing field, the alternating displacements of the electrons constitute a current, a real motion of electricity. But this exposition cannot be complete; for an alternating current continues through a condenser even when the space between the plates is a vacuum, with no electrons present, bound or free, to form a current. We must therefore regard the alternating electric field strength as equivalent to a current In a vacuum it is the whole displacement current, while in a material medium it and the bound-electron current combine to form the displacement current.

The important question about a displacement current is whether, like a conduction or convection current, it has a magnetic field, for it is chiefly by means of its magnetic field that a current is able to do work. Maxwell, a Scottish physicist (1831–1879) who was the first to form clear ideas about these questions, assumed that a displacement current, even in a vacuum, is accompanied by a magnetic field at right angles to the current; and his theory, worked out mathematically, suggested to him the existence of electromagnetic waves and gave an explanation of light as consisting of such waves This led later to the discovery of electromagnetic waves by Hertz, and still later to their application in radio or wireless telegraphy and telephony. An experimental demonstration of the magnetic field produced by a displacement current is given later (§522).

518. Electric Oscillations.—In the preceding few pages we have been considering alternating currents of 25 or 60 cycles per second frequency, produced by alternating current generators. Alternating currents of thousands or millions of cycles per second are not merely possible but are easily produced, though the method of production is different. Apparently no such possibilities were thought of until 1842 when Joseph Henry, American physicist (1797–1878), while studying the magnetization of steel needles by discharges from a Leyden jar (§391), noted that the needles were sometimes magnetized in one direction and sometimes in the opposite direction, though the direction of discharge was the same. This was contrary to what had been believed up to that time, namely that the discharge of a condenser was always a rush of electricity in one direction, from one plate to the other. Henry's conclusion was that each discharge was in reality of an alternating nature.

519. Period of Oscillation.—The subject was taken up theoretically in 1853 by William Thomson (Lord Kelvin), Scottish physicist

(1824–1907), who found for the period τ of oscillation of a circuit containing a condenser

$$\tau = 2\pi\left(\frac{1}{LC} - \frac{R^2}{4L^2}\right)^{-\frac{1}{2}}$$

where C means the capacity of the condenser and L and R are the self-inductance and the resistance of the circuit. If $R^2/4L^2$ is greater than $1/LC$, the expression for τ is evidently imaginary, and in this case there are no alternations, only a discharge in one direction (Fig. 428b). If $R^2/4L^2$ is so small compared with $1/LC$ that it may be neglected

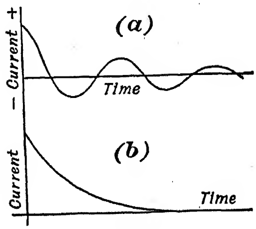


FIG. 428.

$$\tau = 2\pi\sqrt{LC} = \frac{1}{f}$$

where f is the frequency of these *electric oscillations*, as alternations of this high frequency are called. These predictions of theory were confirmed a few years later by photographing the reflection of the spark from a discharging jar in a rapidly rotating mirror (Fig. 429). The frequency of such oscillations is of the order of several millions per second.

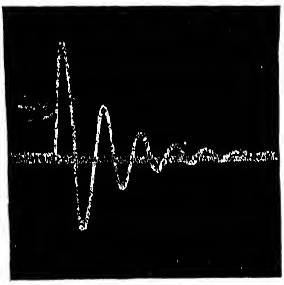


FIG. 429.

The similarity of the formula $2\pi\sqrt{LC}$ for the period of an electric oscillation and the formula for the period of an elastic vibration, such as the $2\pi\sqrt{I/K}$ of §124, is not mere chance. In both cases the oscillations are due to the combined effects of *inertia* and *elasticity*. I in the mechanical oscillations represents inertia, and L in the electric oscillations is electromagnetic inertia (§492). K is a measure of mechanical elasticity, which is stress/strain. C, which is equal to Q/V, is of the nature of strain/stress, since Q is an electric displacement (§517). Hence $1/C$ corresponds to K.

520. Energy Changes in Oscillations.—In the oscillations of a circuit containing a condenser and an inductance, periodic transformations of energy take place. The energy of a charge in a condenser is electrostatic energy, which is evidently a form of potential energy. The energy of a current i in an inductance L

may be called magnetic energy, since the coil has the properties of a magnet; but it may also be regarded as kinetic energy, its magnitude being $\frac{1}{2}Li^2$, which corresponds to $\frac{1}{2}mv^2$ for the kinetic energy of a body in motion. It is instructive to compare these invisible transformations of energy in an oscillating circuit with those in the visible case of a vibrating body, for example a body M (Fig. 430), held between two stretched springs and supported on a horizontal surface with very little friction. In Fig. 430a M has potential energy and C has electrostatic energy; in b M has kinetic energy and L has magnetic energy; in c M has potential energy and C has electrostatic energy; in d M has kinetic energy and L has magnetic energy. Resistance in the oscillating circuit corresponds to friction between M and the supporting surface. Both produce heat at the expense of the energy of vibration; and the vibrations die down, unless additional energy is supplied to keep them up.

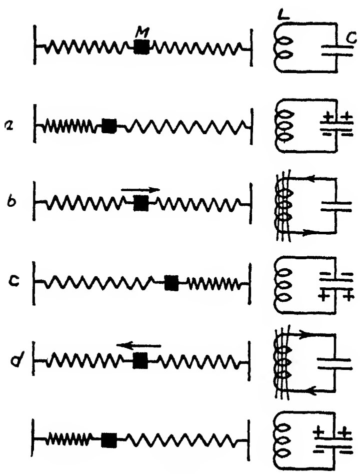


FIG 430.

521. Illustrations of Damped Oscillations.—In the arrangement illustrated by Fig. 431 the primary of a step-up transformer T is connected to a 110-volt, 60-cycle, alternating generator, and the secondary to a condenser C in parallel with a coil L and a spark-gap S. Both the capacity of the condenser and the inductance of the coil are small, so that the natural frequency $1/(2\pi\sqrt{LC})$ of the circuit consisting of C and L, without any gap, is high, about a million cycles per second. The secondary produces a peak voltage of something like 50,000 volts.

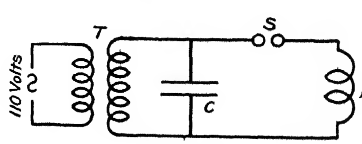


FIG 431.

The condenser is charged, alternately positive and negative, 60 times per second. When the potential difference at S is sufficiently high, a spark passes. This ionizes the air, that is, it disrupts molecules into charged atoms, and the resistance at the spark-gap

is thereby reduced to a very low value. The condenser is then able to discharge with oscillations; but the oscillations are quickly damped out, each group of them occupying only a very small fraction of a cycle of the alternating generator. This occurs about as often as the voltage from the transformer reaches a peak, or about 120 times per second. There are no oscillating discharges from the condenser through the secondary of the transformer, although its resistance is low, because the inductance is too high.

If an incandescent lamp, short-circuited by a single loop of wire (Fig. 432) is placed in the circuit of the condenser C and the inductance L, the oscillating discharge from the condenser passes through the lamp and lights it up, instead of going through the loop of much lower resistance than the lamp. The reason is that, at these high frequencies, the *resistances* of lamp and loop have practically no influence on how the current distributes itself between them. It is the *reactance* that counts (§506) and this is much greater for the loop, owing to its large area, than for the lamp.

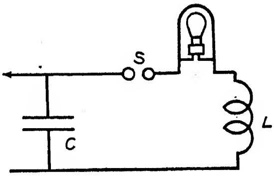


FIG. 432.

522. Electromagnetic Radiation.—In the experiments just described oscillations take place in a circuit of wires connected to a condenser. These surgings of electricity forward and backward constitute wave motions in the circuit, just as the vibrations of a piano wire are waves traveling to and fro along the wire. In both cases the vibrations are rapidly damped out; but, so long as they last, they are stationary waves (§253). While we have spoken of electric oscillations as motions of electrons, currents of electrons are always accompanied by magnetic fields in the space surrounding the circuit. We can picture the magnetic vibrations that accompany the electrical vibrations as rapid expansions and contractions of circular magnetic lines of force, centered on the circuit. These magnetic oscillations extend, more or less, into all the space surrounding the oscillating circuit and can produce other currents in neighboring circuits. Two experiments will illustrate this effect.

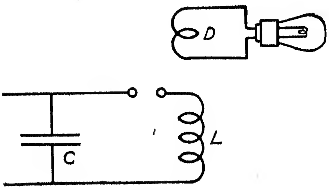


FIG. 433.

A coil D, consisting of a few turns of wire, is held above L (Fig. 433) and is in circuit with an incandescent lamp. When oscillations take place in the circuit of C and L, the lamp lights up. If D is turned through 90° the lamp goes out. The lamp does not light up at all if L is connected directly to the 110-volt circuit, without any condenser C being present. The effect with the oscillating discharges from the condenser is due to the enormously high frequency of the oscillations. The principle of this experiment is often used in "out-gassing" some metal part of apparatus in a vacuum. Oscillations in an external circuit produce eddy currents in the metal and heat it to a red heat, so that it emits occluded gases, which can then be removed by the pump.

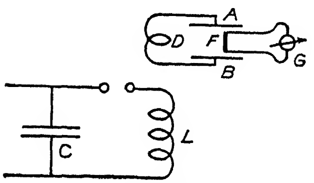


FIG 434.

In the second experiment the arrangement is varied by replacing the lamp by a pair of conducting plates A and B (Fig. 434). Between these plates displacement currents of high frequency are produced, even if the plates are in a vacuum. These currents, acting through their magnetic fields, cause currents in a short wire F. The wire is connected to a thermoelectric form of galvanometer G, which responds to the currents, because it is affected only by the heat produced by the currents, and this does not depend on their directions.

523. Electromagnetic Waves.—Surrounding any alternating current circuit there is an alternating magnetic field associated with the current, and this alternating magnetic field can be pictured as expanding and contracting circular magnetic lines of force (§424). But this is not a complete picture of the alternating field. The current itself is a flow of electrons. Electric lines of force start from or end on electric charges. Here also we are using the graphic method of lines of force as a device for picturing the electric fields that are always associated with charges of electricity. When an alternating current is passing in the circuit, there is at any moment an excess of electrons, negative charges, at some part of the circuit and at some other part there is a deficiency of electrons, which means the presence of positive charges. Between the positive charges and the negative charges there are electric lines of force, curving around through the medium; and, as the alternating current flows, these curved electric lines move out from and back toward

the conducting wire. So we see that near the circuit there is an alternating magnetic field accompanied by an alternating electric field, the two being at right angles. For alternating currents of ordinary frequencies the alternating electromagnetic field near the circuit involves practically no loss of energy of the current; the outflow in one half pulsation is followed by a return flow in the next half pulsation. But when the frequency becomes great the result is quite different.

When the frequency of the alternations runs up to a million or more per second, less and less of the outflowing energy gets back to the circuit. It is thrown off as radiation into space. It can be shown that this radiant energy emitted is proportional to the fourth power of the frequency, that is, doubling the frequency multiplies the radiation by 16. The radiation so emitted consists of waves, electromagnetic waves or radio waves, as they are now called. The wave-front is a surface perpendicular to the direction in which the waves are traveling. What is taking place in the wave-front is an electromagnetic vibration, consisting of an alternating magnetic intensity and an alternating electric intensity. These two vibrations are at right angles to each other and they are in the same phase, that is, they become maxima together and zero together.

This remarkable theory of electromagnetic waves was worked out mathematically by Maxwell and published in 1864. He went much further and explained all light as consisting of such electromagnetic waves of very short wave-lengths; for he found that these waves would have the fundamental properties of reflection, refraction, etc., characteristic of light. But he did not publish this theory until he had ascertained that it would stand one searching test, namely, that the velocity of the waves in any medium should be the known velocity of light in the medium. This test it stood remarkably well, with some exceptions that were accounted for later. Maxwell's formula for the velocity v was $c/\sqrt{K\mu}$, where c is the number of e.s.u. of charge in one e.m.u. of charge, K the dielectric constant of the medium in e.s.u. and μ its permeability in e.m.u. For space and approximately for air $K = 1$ and $\mu = 1$, so that $v = c$. Now we have already stated that $c = 3 \times 10^{10}$ cm. per second and this is also the velocity of light.

While Maxwell's theory was accepted by most physicists as soon as it was understood, it was over 20 years before a method of generating and testing electromagnetic waves was discovered.

524. Hertz's Experiments.—A method of generating and detecting electromagnetic waves was devised in 1888 by Heinrich Hertz, German physicist (1857–1894). One of the simplest of his arrangements is illustrated in Fig. 435. The *oscillator* for generating the waves consisted of two spheres S and S' movable along conducting rods A and B between which there was a spark gap P. A and B were connected to the secondary of an induction coil. When a spark passed at P the air was ionized and became a conductor, so that S and S' were connected by conductors of low resistance. S and S' served, in effect, as the plates of a condenser of very small capacity. The dielectric between the plates of this condenser was

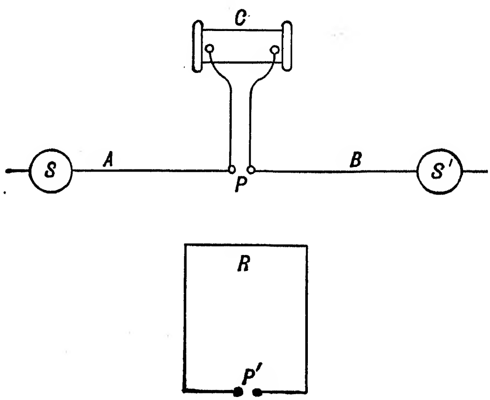


FIG. 435.

the whole of surrounding space. The capacity of the condenser could also be varied by moving S and S' along the rods.

Before a spark passed at P, S and S' were at different potentials. When the spark passed, oscillations started up, the current in the oscillating circuit being a conduction current in A and B and in the gap, and a displacement current in surrounding space. As the self-inductance of the circuit and the capacity were both very small, the frequency $1/(2\pi\sqrt{LC})$ was very high.

As a *detector* a loop of wire R with a spark gap P' was used. This was a circuit capable itself of oscillating and therefore able to take up oscillations of the same frequency as those of its own natural vibrations. When used as a receiver it was, to use an acoustical analogy, a resonator, like a tube that resonates to a tuning fork (§586). The purpose of having S and S' movable along the rods was

to be able to tune the oscillator to exactly the frequency of the receiver.

With apparatus of this general type, Hertz was able to repeat, with electromagnetic waves, the ordinary optical experiments of reflection from metal plates, refraction by a block of paraffin, and the interference of waves, forming nodes and loops, and even to construct a lens of pitch that focussed the waves as a glass lens focusses light. He also found a rough value for the velocity of the waves that agreed with the known velocity of light. In fact Hertz's experiments confirmed the whole of Maxwell's theory of electromagnetic waves and were promptly followed by an application of it in a field of immense practical importance.

525. Radio Telegraphy.—The distance between oscillator or transmitter and detector or receiver in Hertz's experiments extended only to a few hundred

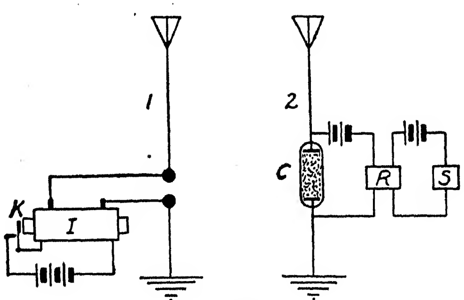


FIG. 436.

feet. Two things were needed to make it possible for this arrangement to develop into a practical system of long-distance telegraphy—a transmitter that would produce waves of greater intensity and more widely extended wave-front and a receiver of greater sensitiveness. In 1895 Marconi, an Italian student twenty years old, began experiments with different devices; by 1899 he was sending signals across the English Channel and in 1901 across the Atlantic.

In Marconi's simplest form of transmitter one of Hertz's spheres was replaced by a long vertical wire or antenna and the other by an earth connection (Fig. 436). This arrangement produced "half-free" waves, their ends being on, and traveling along, the ground. This arrangement was improved, so that more power could be applied, by using the oscillating condenser and inductance arrangement of Fig. 431, the antenna being in series with the secondary of an air transformer of which L was the primary. Many other modifications were introduced from time to time but cannot be considered here.

Ingenious receivers on a wide variety of principles were devised by Marconi and others. All made use of an antenna, and between it and the earth connection there was placed some device that usually depended in some way on a

telephone receiver. One such device that rendered good service for several years consisted of a conducting crystal (galena, carborundum or silicon) touched by a pointed wire. Through such a contact a current could pass in one direction more readily than in the opposite direction. The electromagnetic waves that fell on the antenna produced oscillating currents in the antenna; but only one half of each oscillation went through the crystal, the return half passing through the telephone receiver (Fig. 437). Although because of its more sluggish action the telephone receiver could not follow each single impulse in a train of waves of such high frequency, it did respond to the combined effect of the impulses, all in the same direction, imparted to it by the group of oscillations due to a train of waves. Greatly improved effects were obtained by using a receiver that could be tuned, similar to the transmitter last described (page 463) but reversed in its action. In these ways it was possible to send and receive signals corresponding to the dots and dashes of ordinary telegraphy; but these methods for transmitting and receiving, while remarkable as steps in the evolution of a great art, are now chiefly of historical interest alone, for they have been practically superseded by methods which we shall consider next. To explain these we shall have to begin with a digression into what will seem at first a very remote field of research.

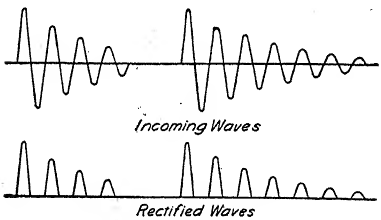


FIG. 437.

526. Emission of Electrons from Metals.—While developing the incandescent lamp, Edison noted the gradual blackening of the inside of the bulb by evaporation emitted by the carbon filament. When a plate P was inserted in the evacuated bulb and connected through a galvanometer G (Fig. 438) to a terminal of the battery that heated the bulb, there was no current if the connection was to the negative side of the battery, but there was a current from the battery to the plate when the connection was to the positive side. This curious observation, made in 1883, was unexplained until, about 1897, the existence and properties of electrons were revealed by the study of the discharge of electricity in gases. The *Edison effect* was then found to be a stream of electrons from the heated filament to the plate, and extensive studies showed that all incandescent metals emit electrons, though in different amounts and at rates dependent on the temperature.

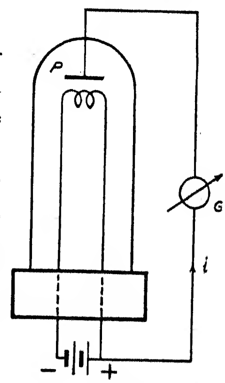


FIG. 438.

When a metal is heated, some of the conduction electrons have velocities much above the average and are able to escape through the surface, in spite of the forces of attraction that the metal exerts on them. In escaping they do work against these forces and their energy is thereby reduced. The escaping electrons remain, for the most part, as an electron cloud, or *space-charge* as it is called, close to the surface of the metal. They can be drawn away by an electrical field directed toward the metal, but not by one in the opposite direction.

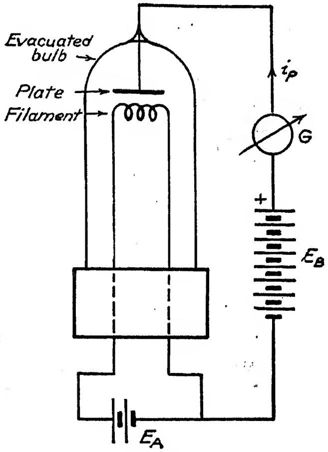


FIG. 439.

In the form of apparatus usually employed for the study of this electron emission (Fig. 439), the battery for heating the filament is always called the "A" battery and the battery connected to the plate through the galvanometer is called the "B" battery. Figure 440 shows, in a general way, the nature of the results with a filament of a pure metal, such as tungsten or platinum. For any temperature T (absolute) of the metal, there is some value of the current above which it does not increase as the plate voltage is raised. This *saturation* current, which is greater the higher T is, occurs when the electrons are being drawn away from the space charge as rapidly as they can be emitted at that temperature. For an impure metal, and especially if there are impurities on the surface, the results are somewhat different. For example, a coating of thorium oxide on the surface of a tungsten filament causes a great increase in the emission of electrons at a given temperature. Moreover the current is very large, and it continues to rise with rise of temperature or plate potential.

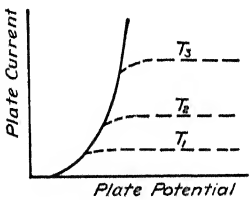


FIG. 440.

It will now be seen that this *two-electrode tube* or *kenotron*, as it is called, is a kind of electrical valve or rectifier, by which alternating currents in an antenna can be turned into a succession of pulses in one direction, that is, a direct current. But this form of

receiving tube has now been mostly superseded by one that has a third electrode added to the other two.

It will be convenient to retain the terms "A" battery and "B" battery, though in practice batteries are usually replaced by connections to a direct current generator or to an alternating current generator through a rectifier.

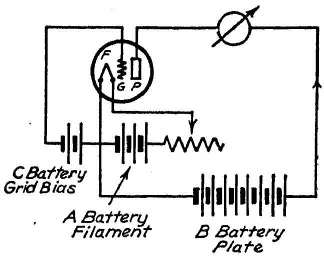


FIG. 441.

527. The Three-electrode Tube. The third electrode added, in the three-electrode tube, to the filament F and plate P of the two-electrode tube, is a *grid*, G, of wire gauze or even a mere spiral of wire surrounding the filament F. It is of fine wire and so widely spaced that it does not directly obstruct, to any considerable extent, the flow of electrons from the space-charge to the plate. Its general purpose may be regarded as *the control of the space-charge.* For example, if it is at a positive potential relatively to the filament it assists the flow of electrons from the space-charge to the plate; and if it is negative relatively to the filament it retards the flow. Moreover it is more effective than the plate in controlling the flow, because it is nearer the filament. Thus the plate current can be changed from zero to some maximum by small changes of the potential of the grid. Figure 442 illustrates in a general way the relation between grid potential and plate current for different values of the plate voltage. These curves are called *characteristic curves* of the tube. By following one of them, say that corresponding to a plate voltage of 50 volts, it is seen how a change of a few volts in the grid potential changes the plate current from zero to a relatively large value, as if the plate voltage had been changed by a much larger amount, though the grid current is always very small compared with the plate current. It will be seen that one part of a characteristic curve is practically a straight line, so that changes of grid voltage and plate current are, for this part, in the same proportion.

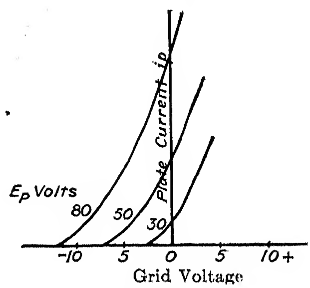


FIG 442.

How this device of a third electrode is applied will be seen by considering the three-electrode tube as (1) an amplifier, (2) an oscillator, (3) a detector.

528. Three-electrode Tube as Amplifier.—An amplifier, as the term is used here, is an automatic device for magnifying electric oscillations by throwing in additional energy from some source,

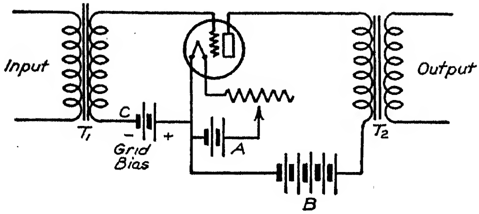


FIG. 443.

which, in the present case, is the "B" battery. Amplifiers can be used at receiving stations or anywhere on the transmission line and also to work loud speakers, and several can be used in series. One such case will illustrate the general principle used.

In Fig. 443 T_1 is an air-transformer. In the primary are the oscillations to be amplified and similar oscillations are induced in the secondary. T_2 is another air-transformer for sending the amplified oscillations out on the line. An essential requirement for telephony or *audio*, as it is called, is that the *form* of the oscillation curve shall not be altered appreciably by the amplification. Transformers without iron cores are used because time-lag in the magnetization of iron would distort the oscillation curve. Fig. 444 shows, in a diagrammatic way, how oscillations of small amplitude in the secondary of T_1 are changed into oscillations of large amplitude in the primary of T_2. The curve is the characteristic curve of the tube at the particular plate potential used. The variations of potential in the input oscillations are impressed on the grid, and the grid is also given an additional, constant, negative potential

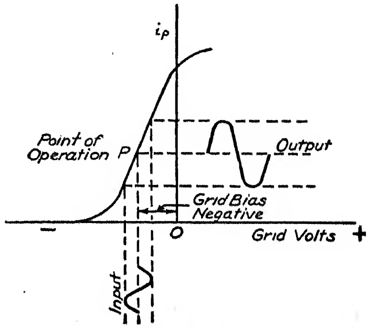


FIG. 444.

or "bias" by the "C" battery, so that the operation of the tube is represented by the *straight part* of the curve. Thus variations in the output voltage, which are proportional to variations in the plate current, are proportional to variations in the grid voltage, which again are proportional to variations in the input voltage. Hence the amplification is free from distortion of the form of oscillation.

The amount of amplification by a three-electrode tube is expressed by its *amplification factor*, which is defined as the ratio of a small change of plate voltage to the small change of grid voltage that would produce the same change of plate current. A ten-fold amplification by a single tube is something like the common practice, but, for special purposes, much higher amplification is possible.

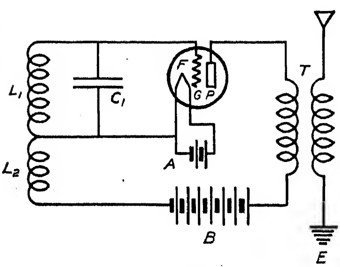


FIG. 445.

529. Three-Electrode Tube as Oscillator.—While the three-electrode tube was originally devised for amplifying, it has been developed into a highly effective oscillator for generating electromagnetic waves. Fig. 445 illustrates, in a general way, the method used. The circuit L_1C_1 is an oscillating circuit of high natural frequency. Any disturbance of it, for example by connecting it to a direct current circuit, will tend to throw it into oscillations, but these would quickly die down if they were not kept up by some regenerative action. The oscillations produce changes in the potential of the grid G, and amplified oscillations are produced in the circuit of the plate, the "B" battery and the filament. Now this circuit also contains an inductance L_2, and the changing magnetic flux of L_2 passes through L_1. If the two circuits are properly tuned, by adjusting C_1, L_1 and L_2, so as to work in unison, energy from the "B" battery will be continuously poured into the oscillations of the plate circuit, and these, acting through a transformer T, will start strong oscillations in the radiating antenna, their frequency being practically that of the L_1C_1 circuit.

Another type of oscillator, of exceptional stability as regards frequency, depends on the piezo-electric properties of crystals (§398). Quartz is especially suitable for the purpose. When a disk of it, properly cut from a crystal, is mounted between two metal plates, like a dielectric in a small condenser, it is set into

vibration, expanding and contracting, by an alternating voltage applied to the plates; and, when it is vibrating, alternating charges are developed on its surfaces. The effects are greatly increased if there is resonance between the applied alternating voltage and the natural vibrations of the quartz. The frequency of the vibrations of the quartz is determined by its dimensions and the elastic properties of the material and is very constant if the temperature is kept constant.

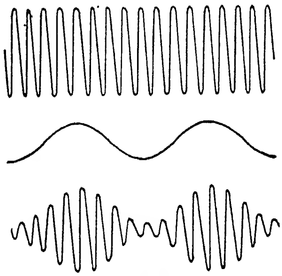


FIG. 446.

The electromagnetic waves sent out by the antenna have a frequency of some millions of cycles per second, called *radio frequency*. They could not be detected by a telephone receiver, for it does not answer to a frequency above a few thousand cycles per second, which is called *audio frequency*. The radio waves are therefore *modulated* so as to serve as "carriers" of audio waves, that is, their amplitude is made to vary according to the pattern of the audio waves. In Fig. 446 the upper curve represents radio waves, the middle curve represents audio waves and the lower curve shows the effect of superposing the two. For example, if the radio frequency is a million and the audio frequency a thousand cycles per

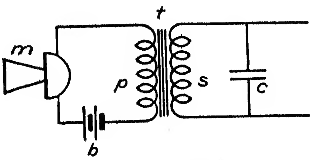


FIG. 447*a*.

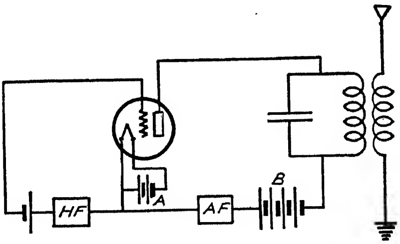


FIG. 447*b*.

second, in the resultant waves there are a thousand radio waves to each audio wave.

One form of *modulator*, or device for producing a periodic variation of the amplitude of the radio waves, is shown in Fig. 447*a*. It consists of a microphone *m* in circuit with a battery *b* and the primary *p* of a small transformer *t* (which in this case may have an iron core), the secondary *s* being formed from part of the grid circuit.

In parallel with s is a condenser c forming a "by-pass" through which the radio waves pass freely as displacement currents, while they are modulated, as regards amplitude, by the superimposed waves from the microphone. Figure 447b shows in a general way where the high frequency oscillator (H.F.) and the audio frequency

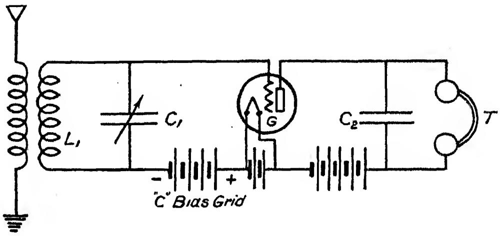


FIG. 448a.

modulator (A.F.) are applied. This is, however, only the briefest possible sketch of arrangements that in practice are very complex.

530. Three-electrode Tube as Receiver.—Fig. 448a indicates a way in which the three-electrode tube is used as a receiver or detector. The circuit C_1 L_1 is tuned to the frequency of the radio waves received by the antenna, and its oscillations of potential are impressed on the grid G. In this case the grid is given such a negative potential by the "C" battery that the operation of the tube is represented by the lower bend of the characteristic curve corresponding to the plate potential (Fig. 448b). The result is the elimination of about half of each incoming radio wave, leaving the portion included under "outgoing wave" to pass to the telephone receiver T and a condenser C_2 in parallel with T.

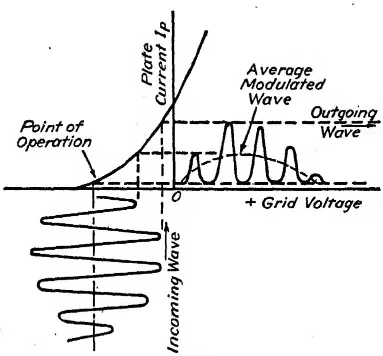


FIG. 448b.

Taken together, T and C_2 constitute a demodulator, that is, it reverses the action of the modulator (Fig. 447a). The high frequency pulses, which can pass through C_2 as displacement currents, but not through T, are separated from the modulating wave, which is of a frequency to which the telephone receiver can respond. Thus

the original waves impressed on the oscillator by the modulating microphone are reproduced in the telephone receiver T.

531. The Thyratron Tube.—The thyratron is essentially a three-electrode tube containing a very small amount of mercury vapor. It thus combines to some extent the properties of a three-electrode vacuum tube and a gas discharge tube.

Consider a three-electrode tube in which the grid potential is sufficiently negative to keep any electrons from reaching the plate. Around the filament there is a negative space charge that prevents any plate current. If the grid potential is made slightly less negative so that a few electrons would go over to the plate, then, in the thyratron, these few electrons ionize mercury atoms and produce positive ions, which move to the grid and neutralise the negative charge around it. The control of the grid is then lost and the current rises to a value limited only by the filament emission or the circuit resistance.

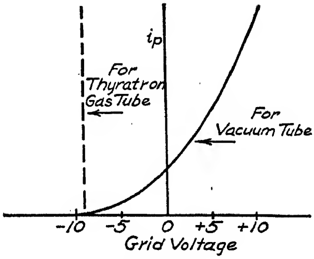


FIG. 449.

The solid line of Fig. 449 shows a *characteristic curve* for a three-electrode vacuum tube and the vertical dotted line is the characteristic curve for a thyratron. For this thyratron -9 volts on the grid prevents any discharge at all, while a small change, perhaps to -8.9 volts, allows a large current to pass, but a change back to -20 will not stop the discharge. The only way to stop it is to remove the anode potential or reduce it to such a value that the electrons no longer have sufficient energy to produce ionization, so that positive ions and electrons combine to form neutral atoms, the process requiring perhaps only 10^{-5} second, whereupon the grid regains control. The ability of a thyratron tube to control the flow of a large current in any way desired, by the expenditure of a small amount of power in the grid circuit, is used in many ways, but the details belong to electrical engineering.

532. Reflection of Electromagnetic Waves.—Light consists of electromagnetic waves of wave-lengths that are only very small fractions of a millimeter. The best reflectors of light are polished metals, the polish being necessary because of the shortness of the wave-length. With longer electromagnetic waves any large conducting surface, however rough, will suffice. Metals are the best reflectors because they are conductors. Electromagnetic waves falling on them produce currents and these give rise to reflected waves. Now high up in the atmosphere there is a large conducting layer, sometimes referred to as the Kennelly-Heaviside layer but now more usually called the *ionosphere*. It consists of gases at very low pressure that are much more highly ionized than those

at the earth's surface and therefore good conductors, the ionization being due to cosmic rays or to ultra-violet rays from the sun.

Strange as it may seem, it is the existence of the ionosphere surrounding the earth and its lower atmosphere that makes long distance radio communication possible. Radio waves sent out by the transmitting station are reflected back by the ionosphere and reach the receiving station. For distances between the two stations up to 50 or 100 miles, waves traveling close to the ground might suffice for communication; but for distances of a thousand miles or more ground waves play practically no part, and communication is by waves reflected by the ionosphere. The enclosing sheath of the ionosphere is so effective as a reflector that signals have been received after traveling completely around the earth. While its height can only be stated very roughly as about 200 miles, there is evidence that its lower boundary rises considerably at night and descends again with the approach of sunrise, which shows that the ionization is due primarily to radiation of some kind from the sun.

While we have referred to only one ionized layer in the atmosphere, it is now known that there are several others, sufficiently distinct and definite to be studied by their effects on radio transmission. They are closer to the earth and are, at times, more effective than the Kennelly-Heaviside layer in making long-distance radio possible. This new physics of the ionosphere has now become a very extensive subject and rapid advances are being made in the study of it.

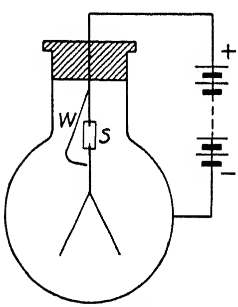


FIG. 450.

CONDUCTION IN GASES

533. Gases as Conductors.—At the beginning of the twentieth century the stage had been set for great advances in physics. Faraday and Maxwell had found relationships between electricity and light; and Hertz, by his discovery of electromagnetic waves, had laid the foundation for modern radio. That electricity might, like matter, consist of indivisible particles had been suggested by Helmholtz, as a consequence of Faraday's work on electrolysis. But nothing whatever was known as to the inner structure of atoms; the first clue to this came from the study of discharges of electricity in gases.

Normal dry air is nearly but not quite a perfect non-conductor. The apparatus illustrated by Fig. 450 was designed to test whether clean dry air conducts at all. The gold leaves of an electroscope are suspended from a sulphur block S, which is fused to a rod that passes through the ebonite cover of the flask. The flask is silvered on the inside. A thin iron wire W is pulled to the right by a magnet (outside the flask) so as to connect the gold leaves to the high-potential battery for a moment, after which it falls back. If any convergence of the charged leaves, showing a loss of charge, takes place, it cannot be by leakage through the sulphur, for the battery would then tend to charge the leaves through the sulphur. The loss of charge could only be through the air to the silvering on the flask. It was found that the leaves did converge very slowly, thus showing that air conducts to some slight extent.

Extensive studies of the subject have shown that air always contains a few charged molecules (or atoms) called *ions*, both positive and negative. Under normal conditions the number of pairs of ions in air is very small. A cubic centimeter of air contains about 3×10^{19} molecules and of these only a few thousand are ionized. When the air is in an electric field, the positive ions are drawn to the negative electrode and the negative ones to the positive electrode.

Ionization of a gas, or the breaking of neutral molecules into charged parts, can be produced by a variety of causes, such as X-rays, radioactive substances, cosmic rays, ultra-violet light and combustion, all of which we shall consider presently. Some radioactivity and cosmic rays are always present. The process of ionization is the ejection of electrons from neutral molecules, thus leaving them positive ions, and the capture of electrons by other neutral molecules, thus forming negative ions.

534. Current through an Ionized Gas.—In the case of a solid or a liquid, the relation between an e.m.f. applied to a body and the current through it is expressed by Ohm's law. To find what the relation is in the case of an ionized gas, let us suppose that the gas is being ionized by a beam of X-rays (Fig. 451) and that in it there are two parallel plates that are kept at different potentials and are in series with a battery and a galvanometer or ammeter. In the gas there is a stream of positive ions toward the cathode plate and a stream of negative ions to the anode plate. The current through the galvanometer is equal to the sum of the two streams in the gas.

If the X-ray beam is constant and the positions and sizes of the plates are not varied, the current depends only on the voltage applied. It is found that for very small voltages Ohm's law holds, the current being proportional to the voltage, as represented by OM in Fig. 452. The ions do not all reach the plates, for if they did the current would not increase with the voltage. The ions start, but many of them are stopped by meeting other ions and recombination takes place. When the voltage is raised sufficiently, a condition represented by NP is reached, the current remaining practically constant. Nearly all the ions reach the electrodes before recombination takes place, and the strength of the current gives a direct measure of the number of ions produced per second by the X-rays. At still higher voltages the current increases rapidly, as represented by PS. The ions acquire such large velocities that, when they collide with neutral molecules, they produce additional ionization and the process is cumulative. Finally the rush and clash of ions and molecules is so great that the electronic systems of many molecules are violently disturbed and light is emitted in the form of a *spark;* the production of heat motions of molecules causes a sudden expansion of air, giving rise to sound; and a large charge passes along some path of maximum ionization.

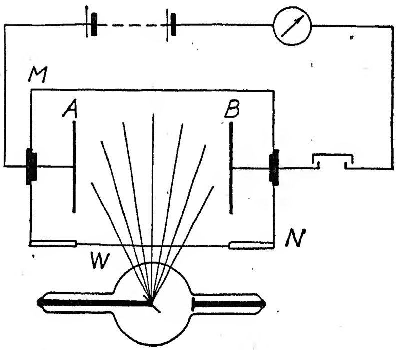


FIG. 451.

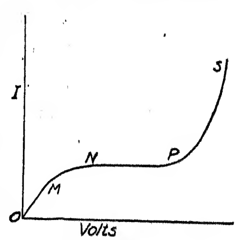


FIG. 452.

535. Effect of Pressure on Sparking.—The potential difference required to produce a spark in air depends on the shape of the electrodes as well as on the pressure of the air. Fig. 453 shows the general results for small spherical electrodes a few centimeters apart.

As the pressure is decreased from atmospheric pressure, the voltage required for sparking decreases, reaches a minimum at about the pressure of 1 mm. of mercury, and then increases as the pressure is further reduced. Finally, at about 10^{-10} of atmosphere pressure, no spark takes place, although there are still about 10^9 molecules per c.c.

There are here two factors that change in opposite directions, as the form of the curve would suggest. At first, as the pressure is reduced the ions ini-

tially present, though reduced in number, can travel a greater distance and acquire greater energy and are therefore more effective in producing additional cumulative ionization. Later, the reduction of the number of ions present becomes the more important factor and finally prevails. The sparking potential is a minimum when the two factors balance.

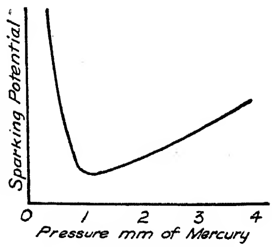


Fig. 453.

In other gases the phenomena of discharge are used for advertising purposes. In an important group, containing argon, neon, helium and mercury vapor, continuous discharges that fill the whole tube take place at pressures of a few cms. of mercury. In these tubes, the red light produced by neon, the pink produced by helium and the blue produced by argon and mercury result from the recombination of electrons with positive ions or from changes in the orbits of electrons in atoms. These processes will be considered more fully later.

536. High-potential Discharges in a Gas.—Striking optical and electrical effects are seen when the discharge from an induction coil or some other source of high potential is sent through air at reduced pressure in a long glass tube with plate electrodes at the ends. In this brief sketch we can mention only a few of the phenomena that seem most important. We shall pass over earlier stages of the reduction of pressure and suppose that the pressure is that of about 0.1 mm. of mercury (Fig. 454*b*), the potential being below sparking potential. On the surface of the cathode there then is a thin band of light called the *cathode glow*. Next to it is a dark space called the *Crookes dark space*, *C*, and beyond that another bright band called the *negative glow*, *G*; then

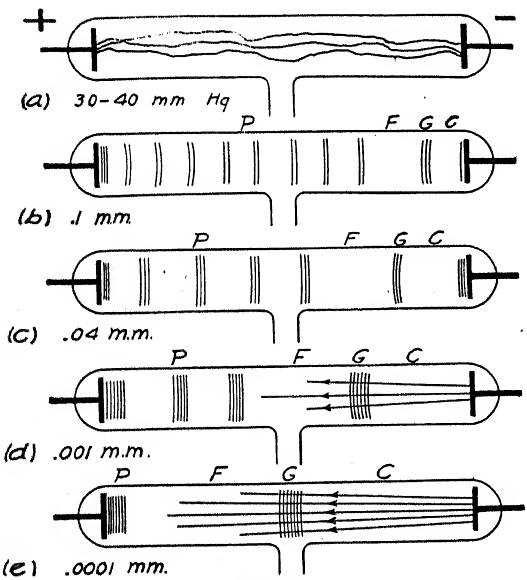


Fig. 454.

comes a second dark space called the *Faraday dark space*, *F*, and the remainder of the tube is filled with *striae* (stripes) of light separated by comparatively dark regions, forming what is called the *positive column*, *P*. To help us in explaining these features of the discharge, we need to know how the potential falls off from the anode to the cathode. This has been found by means of exploring electrodes inserted in the tube. The result is shown in a general way, without minor details, in Fig. 455, from which it will be seen that nearly all the fall of potential occurs in the Crookes dark space *C*, whereas, if there were no ions at all in the tube, the fall from anode to cathode would be uniform.

We can now give some explanation of the striated form of the discharge. Initially there are in the tube heavy positive ions of

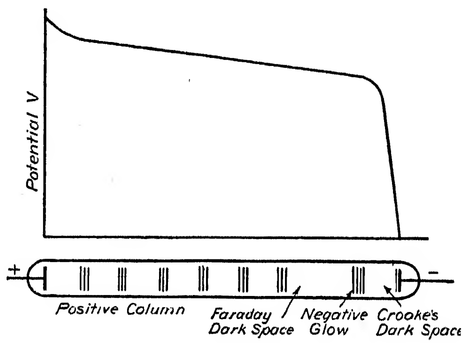


Fig. 455.

molecular dimensions and light negative ions which are electrons. When the electrodes are charged, the heavier positive ions near the cathode rush toward it and knock electrons out of the neutral atoms in the electrode. The electrons, being of small mass and repelled by the cathode, start away from it at a high velocity. This leaves an excess of positive ions near the cathode, thus adding to the field strength there. This accounts for the high cathode fall of potential.

The cathode glow consists of light emitted by the inner electrons in atoms when these inner electrons are disturbed by the ionization of the atoms. The electrons thrown off in the process of ionization travel through the space called the Crookes dark space *C*, without acquiring sufficient speed to produce much further ionization. But beyond *C* they begin to produce ionization, and the negative glow is the result. Having lost energy they again pass through a space, the Faraday dark space, without producing much light, but

farther on they again become active in producing ionization, and a continuation of the process of alternate speeding up and slowing down produces the striae of the positive column.

Imperfect as this brief sketch is, it will afford some assistance in understanding the remarkable results of further decrease of the pressure in the tube.

537. Cathode Rays.—When the pressure in the discharge tube (Fig. 454) is reduced step by step, the Crookes dark space steadily increases so as to occupy more and more of the tube, and at about 0.001 mm. of mercury it fills the tube, all other features of the discharge having disappeared. The position of the anode has no effect on the discharge, and in Fig. 456 it is off to one side. The glass opposite the cathode fluoresces with a yellowish green color, and a small disk of metal, such as the Maltese cross in the figure, casts a shadow on the fluorescing glass, showing that the *cathode rays*, as they are called, travel in straight lines from the cathode. They may, however, be deflected by a magnetic field (Fig. 457).

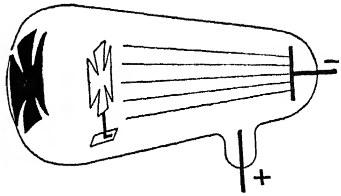


Fig. 456.

These cathode rays consist of what we have from the outset called electrons, and, in fact, it was in this way that the existence and properties of electrons were first discovered. The discovery, one of the most important ever made, cannot be wholly credited

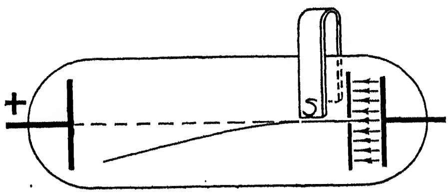


Fig. 457.

to any one man, for many had, by their studies of discharge in gases, prepared the way; but the work of Sir J. J. Thomson, British physicist, on "Cathode Rays" was the most important single contribution.

Thomson determined the ratio of charge to mass and the velocity of electrons in a discharge tube (though he called them cathode rays or corpuscles) by the form of apparatus shown in Fig. 458. Elec-

trons produced in a discharge tube (at the left) passed through a small opening O and then between two plates M and N, where they could be subjected to an electric field between M and N or to a magnetic field perpendicular to the plane of the figure. Either field would deflect the electrons from the path OX to a path such as OZ, ZX being in the direction of the electric field and therefore perpendicular to the magnetic field.

The velocity of the electrons was determined by applying both fields at once and adjusting them, as regards magnitude and direction, so that their effects were opposite and equal and the stream of electrons followed the same path OX as when neither field was applied. Let R be the strength of the electric field in electrostatic units. If the charge of an electron is e in electromagnetic units, it is ec in electrostatic units, where c is the number (3×10^{10}) of

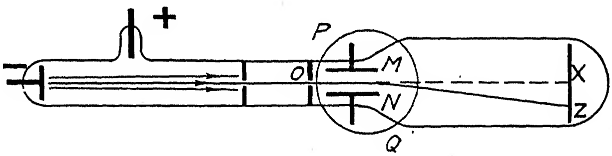


FIG. 458.

electrostatic units of charge in an electromagnetic unit. Then the force in dynes exerted by the electric field on the electron is Rec. To find the force exerted by the magnetic field, we regard the moving electron, in accordance with Rowland's experiment, as equivalent to an electric current. If its velocity is v, it is equivalent magnetically to unit length of a current of ev electromagnetic units (§433) and the force exerted on it by the magnetic field is Hev (assuming the permeability of the field to be unity). Hence, when there is no deflection of the electron stream,

$$Rec = Hev$$

and

$$v = \frac{R}{H} \times 3 \times 10^{10} \text{ cm./sec.}$$

The values of v varied with the voltage. They were between one-thirtieth and one-third of the velocity of light or from 6,000 to 60,000 miles per second.

When the moving electron is subjected to the magnetic field H alone, its path, when it is in the magnetic field, is a short arc of a circle of radius r, and r can be calculated from XZ and the

distance of XZ from MN. Equating the force toward the center of the circle to the force exerted by the magnetic field we get

$$\frac{mv^2}{r} = Hev \qquad \text{and} \qquad \frac{e}{m} = \frac{v}{rH}$$

By this method Thomson found that e/m for a cathode projectile is about 1846 times larger than e/m for a hydrogen ion in electrolysis. This indicated that either e was much larger for the cathode particle than for a hydrogen ion or m was much smaller. From other considerations Thomson adopted the latter view and found evidence for it from a rough determination of e by a separate method. The matter was, however, not settled quite definitely until e, the charge we call an electron, was determined by Millikan's oil drop experiment (§400).

In the *cathode ray oscilloscope* an alternating current, connected to plates M and N as in Fig. 458, produces vibrations of a spot of light, formed by a cathode ray pencil on a phosphorescent screen. The vibrating spot, as seen in a rotating mirror, shows the form of the a.c. curve.

538. Positive Ions.—In a discharge tube there are positive ions moving toward the cathode; and, while they do not give such striking effects as the negative ions or cathode rays yield, they have led to very important information regarding the constitution of matter. But definite information about the positive ions was not obtained by Thomson until 1909, because the investigation was a very difficult one, owing to the high ionization of the residual gas produced by these heavy positive ions.

To isolate and study the positive ions, a cathode that had a very fine tunnel in it parallel to the axis of the discharge tube was used, so that some positive ions shot through it into a second chamber B (Fig. 459). In B it was necessary to keep up a much higher vacuum, by means of a side tube, than the vacuum in the discharge tube, in order to avoid breaking up the stream of positive ions by collisions with neutral molecules. In B the ions passed through an electric field, which exerted a transverse force on them in the direction of the field, and also through a magnetic field in the same direction as the electric field, so that they were subjected to a second transverse force perpendicular to the direction of the first. If the ions had been all similar and of equal velocity, they would have landed at practically the same point on a photographic plate at

the end of the tube. If they were similar but with different velocities they would land on a parabola, the equation of which can be shown to be

$$y^2 = A\frac{ne}{M}z$$

In this e is the charge of an electron, n is the number of electrons

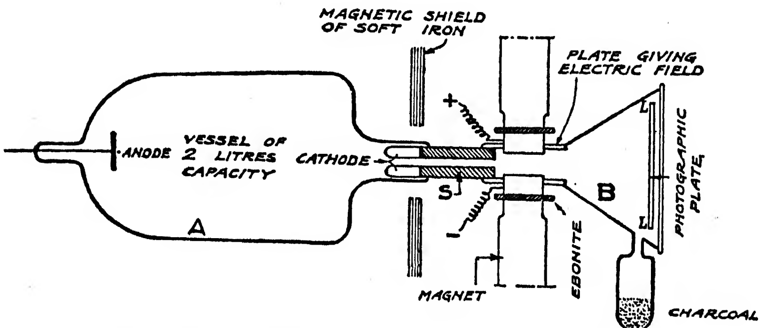


FIG. 459.—(From Castelfranchi's "Recent Advances in Atomic Physics.")

removed from an atom when it was ionized or ne is the charge on the positive ion, M is the mass of an ion, and A is a constant that depends on the dimensions of the tube and the strengths of the fields. Parabolas for $n = 1$, $n = 2$, $\cdots$, were obtained but those for $n = 1$ were the brightest, showing that most ions were singly ionized. Let us now confine attention to parabolas (Fig. 460) for which $n = 1$. Each gave a value of e/M, and, for two ions of different masses, it gave the ratio M/M' of the masses. This method of detecting minute quantities of chemical elements and comparing their atomic weights is much more effective than chemical analysis.

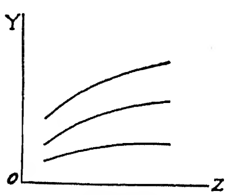


FIG. 460.

539. Isotopes.—The positive ray method of analysis was further developed by Aston in his *mass spectrograph* (Fig. 461). The ions, after passing through the electric field at MN, where they are separated according to their masses and velocities, pass through a magnetic field H, so designed that all of the same mass, whatever their velocities, fall on the same point on a line PQ, the points for different atomic weights being separated along PQ. It was thus found that the atoms of a chemical element are not all of the same atomic weight. For example neon has atoms of atomic weight 20

and others of atomic weight 22, mixed in such proportion as to give the atomic weight 20.2 found by chemical methods. Atoms of the same chemical properties but different atomic weights are called *isotopes*. Most elements have several isotopes; hydrogen has three, of atomic weights 1, 2, 3, oxygen also three, of atomic weights 16, 17, 18, the first in each case being the most numerous, while tin has eleven, with atomic weights between 112 and 124.

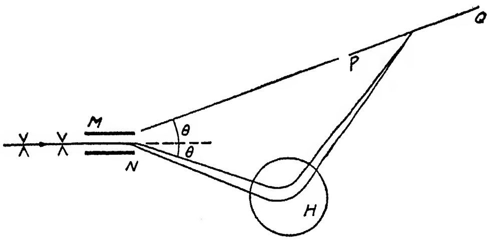


FIG. 461.

This subject is of such great importance to the chemist as well as to the physicist that a great deal of attention has been devoted to it, and better forms of mass spectrographs have been devised.

540. Structure of an Atom.—While the study of positive ions in a discharge tube extended the bounds of knowledge as regards the variety of atoms, it was a use of cathode rays that started Physics on its discoveries regarding the inner structure of an atom. In 1903 Lenard, a German physicist, found that fast cathode rays would pass through thin sheets of metal foil, and this suggested to him that an atom is not a hard sphere, as formerly thought, but consists of a massive center or *nucleus* with electrons surrounding it at comparatively large distances. But electrons are very small tools for probing the interior of an atom, and in 1911 Lord Rutherford bombarded thin gold foil (2×10^{-5} cm thick) with α-particles from radium, each of which has approximately the mass of four hydrogen atoms and carries a double positive charge (§548). They were not deflected by the electrons between which they passed, but many of them were deflected through large angles, as ascertained by observing with a microscope flashes produced by the impacts of the particles on a zinc sulphide fluorescent screen (Fig. 462). It

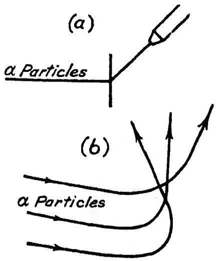


FIG. 462.

was found that the number scattered through any angle could be accounted for by assuming that the positively charged particles were repelled by positive charges constituting the nuclei of atoms. Thus the term *nuclear atom* came into use as descriptive of this picture of the constitution of an atom.

Evidence has accumulated that the "nuclear radius" is of the same order of magnitude as the "radius" of an electron, or about 10^{-13} cm., and small compared with the "radius," 10^{-8} cm., of the whole atom, though the term "radius," used in this connection, does not mean something as definite as when it is applied to a visible body. The arrangement of the electrons in an atom in "shells" has already been referred to in explaining the magnetic and diamagnetic properties of substances (§449). The chemical properties of an atom depend on its outermost shell. If this is filled or complete, the atom is very stable or chemically inactive, as in the cases of helium, neon, argon, etc. If it lacks an electron, the atom readily takes on an electron and becomes negatively charged; if it has only a single electron, the atom readily gives up the lone electron and becomes positively charged. Thus Physics would seem to have supplied a new basis for Chemistry. Yet there is reason to believe that the picture here presented is only a simplified outline of reality.

541. X-rays or Roentgen Rays.—The continued study of the phenomena associated with discharge tubes led, in 1895, to a remarkable discovery. Roentgen, a German physicist (1845–1923) found that a high-vacuum discharge tube emitted a radiation that affected a photographic plate wrapped in black paper and caused a fluorescent screen to glow. Not knowing the nature of the rays, he called them by the non-committal name of X-rays. Later it was shown that they were emitted from the glass walls of the tube as a result of the impact of cathode rays. How this first property of X-rays of *passing through opaque substances* to a greater or less extent, depending on the density of the substance, has proved of service to the surgeon, the dentist and others is too well-known to need description here.

A second property of X-rays, to which we have already referred, is their power of *ionizing* gases (§533). A third property, which includes their photographic action, is the readiness with which they produce brilliant *fluorescence* of various substances, especially of the platinum-barium-cyanide in the fluoroscope used for visual observa-

tion of the internal organs of the body and many other purposes. The *chemical action* of X-rays forms the basis of many applications in medicine. These are only a few of the numerous uses of X-rays.

As to the fundamental nature of these rays there was at first much doubt, for they did not seem to show the properties of reflection, refraction and interference associated with light; but these were discovered later, and there is now no doubt that X-rays are electromagnetic waves of wave-lengths much shorter than ordinary light. One fact that pointed to this conclusion from the beginning was that, like light, they were not deflected by electric or magnetic fields, as cathode rays or positive rays are. The mechanism by which they are emitted by atoms when bombarded by fast electrons or cathode rays is also to some extent understood. The bombarding electron knocks an electron out of the innermost or K-shell into some outer shell. To displace an electron from the K-shell requires more energy than to remove one from an outer shell, since the K-electron is nearer the nucleus and more strongly attracted by it. Another electron from an outer shell falls into the vacancy created in the K-shell, and, in doing so, it gives off energy in the form of X-rays (§731). Everything else, such as the wave-lengths of X-rays, checks up with this explanation from the point of view of energy. But here our knowledge of the mechanism of the process stops short. How an electron dropping from an outer orbit to an inner one emits waves we do not know, and the process is probably more complex than the simple picture of the drop of the electron represents.

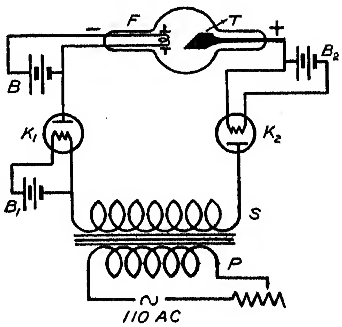


FIG. 463.

542. X-ray Tubes.—The earlier tubes for producing X-rays resembled the one originally used by Roentgen in having a gas at very low pressure through which electricity was discharged. Later, these so-called "gas tubes" were replaced by the Coolidge form of tube (Fig. 463). In it the vacuum is so high that no discharge between cold electrodes would be possible, but electrons are emitted by a hot filament F and are given a high velocity by a strong electric field between F and a positive electrode T, called the target, from which the X-rays are emitted by the impact of the electrons.

F is a small spiral of fine tungsten wire and is heated by the battery B. T is a large block of metal that does not rise much in temperature with the impact of the electrons.

The electric field between F and T is produced by the secondary S of a transformer, the primary P being connected through a resistance to an alternating current circuit. The alternating current produced in S is rectified by two two-electrode tubes or kenetrons K_1 and K_2 (§526). The potential applied to the tube by the kenetrons is a pulsating direct-current potential, and the averaged current from the filament is only a few milliamperes. Even a very small amount of gas in the tube would produce positive ions, which, impinging on F, would quickly burn it out by rapid bombardment. The velocity of the electrons in the X-ray tube is determined by the voltage applied to it. If the mass of the electron is m and the velocity given to it by the field is v, the kinetic energy it acquires is $\frac{1}{2}mv^2$. If its charge is e, in falling through the potential drop V the work done on it by the field is eV. Hence

$$\tfrac{1}{2}mv^2 = eV$$

The velocity v determines the "hardness" or penetrating power of the X-rays. The batteries B, B_1 and B_2, being at high potentials, must be well insulated.

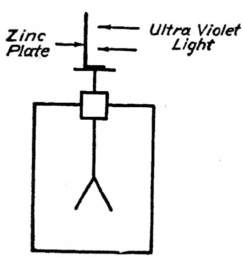


FIG. 464.

543. The Photoelectric Effect.—The property that X-rays have of ejecting electrons from atoms is shared by another form of electromagnetic waves. *Ultraviolet light* is an invisible form of light that is produced, along with visible light, by an arc-light or mercury vapor lamp. It can be studied by means of the fluorescence it causes and by its photographic action. It includes a range of wave-lengths between the longer wave-lengths of ordinary light and the shorter wave-lengths of X-rays. A glass plate that is transparent to visible light is practically opaque to ultraviolet light, whereas quartz freely transmits the ultra-violet.

If a zinc plate is placed on the cap of an electroscope (Fig. 464) and both are charged negatively, a beam of ultraviolet light, thrown on the plate, causes the leaves to collapse; but no such effect is produced if the plate and electroscope are charged positively. This

is called the *photoelectric effect*. It takes place when the plate and electroscope are in a vacuum, into which the ultraviolet light is admitted through a quartz window. Hence the effect is not due to ionization of the air or other gas. Electrons, called *photoelectrons*, are liberated from atoms by the ultraviolet light; but, when the plate is charged positively, they are kept from escaping by the attraction of the positive charge. Many other metals give similar results, but they do not all require ultraviolet light to show the effect. In fact, for some metals, such as potassium, rubidium and caesium, visible light or even infrared light will suffice. For each metal there is a certain minimum frequency of light vibration below which no appreciable photoelectric effect is produced.

544. Photoelectric Cells.—Devices for detecting or measuring light by its photoelectric action have found many applications. In the telegraphic transmission of pictures on developed photographic plates or films, a narrow beam of light sweeps over every part of the plate or film, or "scans" it, and the fraction of it transmitted produces a proportional electric effect by means of a photoelectric cell. The current, amplified and transmitted, is applied to vary a beam of light that reverses the scanning process and reproduces the picture. In *television* use is made of the light reflected in different amounts from different parts of an object or from different parts of an audience or view. The photoelectric effect, using invisible ultraviolet light, can be applied to detect the presence of a burglar near a safe and sound an alarm. For details as to these and other applications special works of reference must be consulted.

545. Quanta of Energy. Photons.—The importance of the photoelectric effect in its practical applications has been equalled by its influence on our fundamental ideas of energy and radiation. The point of interest in this connection was how the number of electrons emitted from a metal and their kinetic energy after emission depend on the intensity of the incident light and its frequency. The intensity of the light is the energy it imparts per unit area per second to a surface at right angles to the direction of the light. The kinetic energy of an electron is not the same for all electrons, but it is found by experiment that, for a given metal and light of a given frequency ν and intensity I, the kinetic energy does not exceed a certain maximum value, say E_m or $\frac{1}{2}mv_m^2$, where v_m is the velocity of the swiftest electrons emitted. Now considerations based on the wave theory of light suggested that E_m should depend

on I and not on ν, but this was not confirmed by experiment. Two important results obtained for any particular metal are: (1) when electrons are being emitted by the action of light, the *number* emitted is proportional to the *intensity* of the light; (2) the *maximum energy* of the emitted electrons depends only on the *frequency* of the light.

These unexpected but clearcut results of experiment were interpreted in a striking way by Einstein in 1905. His explanation, now generally accepted, was that light, at least in this connection, must be regarded as consisting of indivisible packets of energy, called *quanta* of energy or *photons*. The magnitude of a quantum or photon is $h\nu$, where ν is the frequency of the light, as the term is used in connection with waves, and h is Planck's radiation constant (§705). An atom in the metal absorbs a quantum of energy $h\nu$ and becomes unstable, with the result that one of its least firmly bound electrons is thrown off and carries the absorbed energy with it. If the atom is in the surface of the metal, the escaping electron spends part of the energy $h\nu$ in doing work against the forces that bound it in the atom and another part in doing work against the forces at the surface of the metal that restrain free electrons in the metal from escaping. If the total work w is less than $h\nu$ the electron emerges with energy $(h\nu - w)$. This is the maximum energy of an escaping photoelectron, for an electron escaping from an atom that is not at the surface is obstructed in making its way through the layer of atoms between its starting point and the surface. We thus arrive at Einstein's important equation:

$$E_m = h\nu - w$$

In this w is called the *work function* of the photoelectric effect for the particular metal used. If we put $E_m = 0$, we get for the lowest frequency at which photoelectrons are emitted $\nu = w/h$, and this is called the "threshold" frequency.

The *quantum theory* of radiation, suggested by the study of the photoelectric effect, is confirmed by many other phenomena. In its application to light it does not supersede the wave theory, which is necessary to explain interference and diffraction (§676). Both theories are required to explain all phenomena, and they have been combined in a wider generalization, called *wave mechanics*, to which some consideration will be given later (§758).

RADIOACTIVITY. THE CONSTITUENTS OF MATTER

546. The Discovery of Radioactivity.—The discovery in 1895 of X-rays, with their penetrating power and photographic action, suggested to Becquerel, a French physicist, that some forms of matter might emit similar radiation without stimulation. After a wide search, he found, in 1896, that salts of uranium, the heaviest known element, are *radioactive*, that is, they constantly emit radiations that affect a photographic plate and ionize air. His method of testing was to place the substance in an electroscope case and find whether and at what rate it caused discharge of the leaves by ionizing air.

Following up Becquerel's discovery, Pierre and Marie Curie found that the radioactivity of any uranium compound is proportional to the amount of uranium it contains, but that certain ores containing uranium compounds are much more radioactive than the amount of uranium in them would account for. This indicated the presence of something much more radioactive than uranium itself. After three years of arduous chemical analysis and testing by the electroscope, the Curies isolated two highly radioactive elements, which they named *polonium* and *radium*. Later, other radioactive elements were discovered. All are of high atomic weight.

The discovery of radioactivity was, one of the most important discoveries in the history of physics. It was the starting point of most of the recent discoveries regarding the constitution of matter.

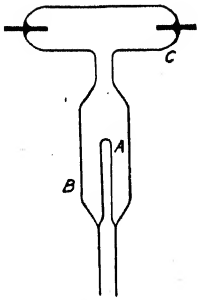


Fig. 465.

547. Particles and Rays Emitted.—The full explanation of the nature of radioactivity came rather slowly. After the pioneer work in France the first distinct advance was made by Rutherford in 1903 at McGill University in Montreal. Preliminary work had indicated the presence of helium in the emission from radium, and the point was tested by the apparatus sketched in Fig. 465. A very thin-walled tube A was sealed into an outer tube B, which was connected to a discharge tube C, B and C being evacuated to a low pressure. Helium gas under pressure was introduced into A and, after some hours, the discharge tube C was used to test for the presence of

helium, but none was found. This showed that there was no leak between A and B. The helium was then removed and an "emanation" or gas, now called *radon*, given off by radium and found to be radioactive, was introduced into A, and, after some time, a test of the discharge tube showed the presence of helium. This proved the presence in the emission from radium of helium ions, capable of passing through the thin walls of A.

Another method of testing was by finding the effect of a magnetic field on a stream of emissions from radium. The general results of this test are shown diagrammatically in Fig. 466. In the emission there are positively charged particles, shown deflected to the left in the figure, negatively charged particles, deflected to the right, and rays that are not deflected at all and are therefore presumably a form of light or electromagnetic waves.

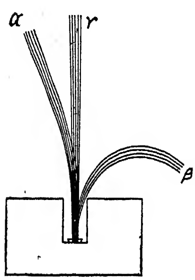


FIG. 466.

548. Radiations from Radioactive Substances. Passing over further details in these very extensive studies, we may now summarize the results as regards the different kinds of emissions from radioactive substances.

(1) *Alpha Particles.*—Each of these has a mass 4, as compared with 16 for the mass of an oxygen atom, and a charge +2, on the basis of −1 as the charge of an electron. They are therefore nuclei of helium atoms. They are emitted with velocities that approach one-twentieth of the velocity of light. The energy of each as emitted by radium is 1.2×10^{-5} erg, or the energy that would be imparted to an electron in falling through 7.66 million volts in a vacuum tube. On account of their relatively great energy, the α-particles are powerful ionizers. Their energy, when in air, decreases rapidly, and they travel only a few centimeters before falling to ordinary velocities. By attracting two electrons that drop into the vacant orbits, each then becomes a neutral helium atom. Alpha particles produce fluorescence in various substances, especially platinum-barium-cyanide, and they affect a photographic plate. Most of the energy of the emissions from radium is energy of α-particles, but their penetrating power is so low that they barely pass through an ordinary sheet of paper.

(2) *Beta Particles.*—The negative particles emitted from radium are electrons of enormous, though different, velocities, sometimes

approaching that of light to within a fraction of one per cent. They also produce ionization and photographic action, though not as intensely as α-particles do, but their much greater speed carries them much farther in air before they cease to ionize and even enables them to pass through aluminum foil several millimeters thick.

(3) *Gamma Rays.*—The third form of radioactive emission consists of electromagnetic waves, that is, it is of the same nature as light and X-rays but of much shorter wave-length than either. Roughly the wave-length of visible light is of the order of 10^{-5} cm., that of X-rays of the order of 10^{-8} cm., while the wave-length of γ-rays is of the order 10^{-10} cm. Their most striking characteristic is their immense penetrating power, in which respect they greatly exceed X-rays, for they can pass through several meters of air or several centimeters of lead.

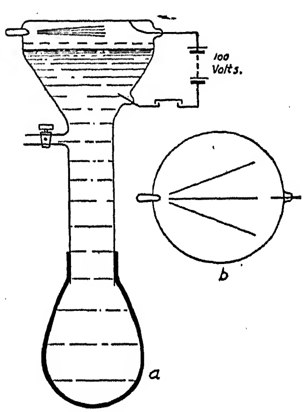


Fig. 467.

549. Wilson's Cloud Chamber.—A strikingly simple device for actually seeing and photographing the paths of α-particles and electrons was devised by C. T. R. Wilson, English physicist, in 1912. It came at once to the aid of the nuclear theory of the atom and has been of immense assistance in many other investigations of these (and other) ultramicroscopic particles.

When air or any gas that is saturated with water vapor is allowed to expand suddenly, it is cooled, and the water vapor is ready to condense on any foreign particles that may be present, especially if they are electrified. In Wilson's cloud chamber air or some gas, carefully cleansed from dust and saturated with water vapor, is in a vertical glass cylinder that is covered by a glass plate and closed below by a column of water that can serve as a piston (Fig. 467). In the simplest form of the apparatus the air is first compressed by squeezing a rubber bulb and is allowed time to become saturated with water vapor. The bulb is then released and the sudden expansion cools the air and vapor. If there are any charged α-particles shooting through the air and vapor, their tracks are shown by a trail of water droplets formed on ions that are produced by impacts of the α-particles on air or water molecules. A strong light thrown

in from the side makes it possible to see and photograph the tracks through the glass cover. When necessary, an electric field can be produced between the top and bottom of the chamber before expansion, to remove any ions lingering in the chamber.

In Wilson's earliest photographs of cloud tracks in 1912 (Fig. 468) not only were sharp lines due to α-particles observed, but some of them were sharply forked, showing the direct impact of α-particles on the nuclei of atoms.

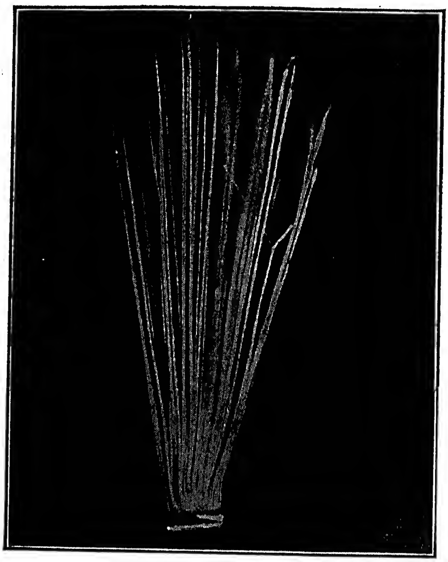


Fig. 468.

550. Radioactive Series and Transformations.—If the nuclear atom is accepted as a correct picture of the constitution of an atom, it is evident that a particle so massive as an α-particle must come from the nucleus of an atom. An atom that has lost part of its nucleus is an atom of a different element. Usually the new atom is itself unstable and, after a longer or shorter time, suffers some further loss, becoming again a different atom, and so on. Finally a stable condition, which is usually an atom of lead, is reached, though it is not always lead of the same atomic weight.

It has been found that there are three such series of disintegrating atoms, starting from different unstable atoms: a uranium-radium series, a thorium series and an actinium series. The uranium-

radium and actinum series originate in uranium or an isotope of uranium. As an example of successive transformations we give the disintegration products of radium:

Element	Radium	Radon	Radium A	RaB	RaC	RaC′	RaD	RaE	RaF	Lead
Atomic number.....	88	86	84	82	83	84	82	83	84	82
Atomic weight......	226	222	218	214	214	214	210	210	210	206
Radiation emitted..	α	α	α	β	β	α	β	β	α	
Half-life period.....	$1580y$	$3.85d$	$3.05m$	$26.8m$	$19.5m$	$10^{-6}s$	$16.5y$	$4.85d$	$136d$	

y, years; *d*, days; *m*, minutes; *s*, seconds.

Atomic number (§540) determines the chemical properties of the atom, and it drops by 2 when an α- particle is emitted. The atomic weight drops by 4 when an α-particle is emitted, but it is not changed appreciably by the emission of a β-particle or γ-rays. The emission of a β-particle raises the atomic number by 1, for it comes from the nucleus and leaves the nucleus with a positive charge larger by 1 unit and therefore ready to capture an additional orbital electron to form a neutral atom. In the example given above RaA, RaC′ and RaF are isotopes, also RaB, RaD and lead are isotopes. The atomic weight of the lead is 206. The actinium series and the thorium series also end in lead, but not this same lead, for the atomic weights are 207 and 208 respectively.

The term "half-life period" means the time required for a large number N of the atoms of the element to be reduced to $\frac{1}{2}N$ by transformation into the next type of atom. Its wide variation shows the wide range of stability of the different temporary atoms. The rate of radioactive decay of one of these temporary elements is not affected, so far as known, by any changes of physical condition, such as extremes of temperature. The internal forces and energies are too great to be affected appreciably by external conditions.

If the number of atoms in a specimen of a radioactive substance is N, the number that disintegrate in a short time dt is proportional to N and to dt and is therefore equal to $aNdt$ where a is a constant that may be called the chance that an atom has of disintegreating in the next second. Hence, if we denote a small *increase* of N by dN (as customary in mathematics) $dN = -aNdt$. From this by integration we get $N = N_0e^{-at}$, where N_0 is the number of atoms at the beginning and N that at the end of an interval t and e is the base of

Napierian logarithms. If we put $N = N_0/2$, the half-life period T is given by $aT = log_e 2$ and $T = 0.693/a$. For example, the rate of emission of α-particles from a gram of uranium I (the principal isotope of uranium) is found to be 1.2×10^4. The number of atoms in a gram of uranium I being 2.5×10^{21}, the value of a is 4.8×10^{-18}. This gives for T 4.6 billion years.

This last result is interesting in its bearing on what is called "the age of the earth" (§328), that is, the time that has elapsed since the surface of the earth became solid. It may be assumed that all the lead of atomic weight 206, so-called uranium lead, existing at present in the earth's crust, has been produced by the disintegration of uranium I. From an estimate of the proportions of uranium I, uranium lead and intermediate products, now present in the earth's crust, and from similar data as regards thorium and thorium lead, it is deduced that the age of the earth is somewhere about three billion years, though the evidence is not yet sufficiently accurate to justify a more precise statement. This result is in good general agreement with results derived by astronomers from two other sources, namely, the peculiarly elongated orbit of the innermost planet Mercury, and the distance to which the moon has receded from the earth as a result of tidal friction.

551. Heat Produced in Radioactivity.—The remarkable nature of these transformation of radioactive substances is shown by the magnitude of the energy involved. This will be clearer if we reduce the energy of all the emissions to heat units, namely calories. The result is that radium in a steady state is giving off energy equivalent to about 100 calories per gram per hour, and this is confirmed by direct experiment. The radium itself remains permanently about 2°C. above its surroundings, owing to some absorption by the radium itself of energy given up by transformation of some of its atoms. It can be calculated that the total amount of energy given up by a gram of radium before it is finally reduced to lead is of the order of 10^{10} calories, which is about the amount of heat produced by burning a ton and a half of coal. Evidently this source of energy would supplant all other, if any considerable amount of radium were available and its radiation could be controlled for use.

The enormous amount of energy apparently stored in the nuclei of atoms has an important bearing on the relations of mass and energy. When the atomic weight of oxygen is taken as 16, that of hydrogen is 1.008. We may take these figures as the relative masses of the nuclei, since the masses of the electrons are negligible. Now if the nucleus of oxygen contains 16 hydrogen nuclei (protons), its mass should apparently be greater than 16. This *mass defect*, as

it is called, ceases to exist, if, following the principle of relativity, we regard the potential energy concentrated in the nucleus as having itself an equivalent mass. When the nucleus explodes the potential energy in the nucleus becomes the kinetic energy of the particles emitted. The question will be considered with more definite figures later (§552).

552. The Transmutation of Elements.—The transformation of one atom into another of different properties was the dream of the alchemists. It is realized in radioactivity, but the process is spontaneous and cannot be directed to the production of any desired form of atom. In 1919 Rutherford succeeded in changing the nuclei of certain atoms into those of other elements by the impact of α-particles. For example, nitrogen bombarded by α-particles emitted protons, which are the nuclei of hydrogen atoms. The reaction that occurs can be represented by an equation that is similar to a chemical equation

$$_2He^4 + {}_7N^{14} \rightarrow {}_8O^{17} + {}_1H^1$$

In this $_2He^4$ stands for the α-particle or helium nucleus of atomic weight 4 and atomic number or nuclear charge 2, and similarly as regards the other symbols, $_8O^{17}$ being one of the isotopes of oxygen. In any such equation the sum of the atomic weights on the two sides must be the same and the sum of the nuclear charges must be the same.

Later it was found that protons are ejected from the nuclei of many of the lighter elements when struck by α-particles. But alchemy of this form was very restricted in its output and extravagant in its use of the element radium. The chance of a collision between an α-particle and a helium nucleus was small, since the cross-sectional area of the helium nucleus is only 10^{-10} of that if the helium atom. In 1932 it was found that swift protons could be used instead of α-particles. For example protons produced in a discharge tube when the voltage applied was 100,000 volts disintegrated the nuclei of one of the lithium isotopes, according to the reaction

$$_3Li^7 + {}_1H^1 \rightarrow {}_2He^4 + {}_2He^4$$

Here the lithium nucleus captures the proton and the resulting structure explodes, with the production of two α-particles.

It is of interest to consider more closely the masses on the two sides of the above equation. The atomic weight of the lithium isotope, as found by the mass spectrograph (§539) is 7.0164, that of

the hydrogen 1.0081 and their sum is 8.0245. The atomic weight of helium being 4.00336, the mass on the right is 8.0067. Thus there is apparently on the right a *mass defect* of 0.0178. The only explanation that can be given for this is based on the principle of relativity. According to it mass and energy are different forms of one entity, an amount of energy E in ergs being equivalent to E/c^2 grams, where c is the velocity of light or 3×10^{10} cm./sec. The energy of the α-particles that are ejected in the reaction can be found from the distance they travel in air before being stopped. It is found to be 0.166×10^{20} ergs for 1.0081 grams of hydrogen in the reaction. Dividing it by 9×10^{20} we get 0.0184, which agrees with the apparent mass deficiency 0.0183 to within the limits of accuracy of the work. This result provides considerable evidence for the relativity principle.

553. The Neutron.—For many years it was believed that there are only two fundamental particles, the electron, charged negatively, and the proton, charged positively. When it was discovered in Germany in 1930 and confirmed in France in 1931 that beryllium, when bombarded by α-particles, gave off something of great penetrating power, the new radiation was regarded as some form of X-rays or γ-rays. Its distinctive character was that it could knock protons out of paraffin (Fig. 469); and by careful analysis of this effect, Chadwick, of Cambridge University, showed in 1932 that the agent must be a particle of nearly the same mass as the proton but *without any charge*. It was therefore called the *neutron*. Because it is uncharged and of large mass, it passes through the cluster of orbital electrons in an atom and is not deflected appreciably, unless it makes a direct hit on the nucleus. It passes more readily through lead than through paraffin because the lead has fewer atoms per unit volume.

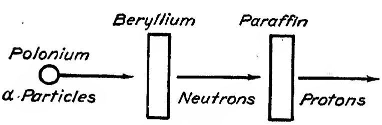


FIG. 469.

Neutrons are very effective in producing nuclear transformations. The following reaction represents the disruption of a nitrogen nucleus by a neutron, giving a beryllium nucleus and an α-particle:

$$_7N^{14} + {_0n^1} \rightarrow {_5Be^{11}} + {_2He^4}$$

A very interesting nuclear reaction, resulting in the production of a neutron, occurs when γ-rays of sufficiently high frequency act on a

deuteron, which is an isotope of hydrogen:

$$_1D^2 + \gamma \rightarrow {}_1H^1 + {}_0n^1$$

This reaction shows that a deuteron is built up of a proton and a neutron.

554. Artificial Radioactivity.—The radioactivity discovered by Becquerel in 1896 was a spontaneous natural process in certain kinds of matter and not in any way controllable by man. Until 1934 no other kind of radioactivity was known, but in that year Joliot and Irene Curie Joliot found that radioactivity could be induced in elements that are usually stable. Aluminum bombarded by α-particles from polonium continued to give off a radiation for some time after the polonium was removed to a distance. A study of the process showed that the reaction of the α-particles on the aluminum was:

$$_{13}Al^{27} + {}_2He^4 \rightarrow {}_{15}P^{30} + {}_0n^1$$

The phosphorus nucleus produced is that of an isotope of ordinary phosphorus. It is inherently unstable and explodes with the ejection of a positron (§555) leaving a stable silicon nucleus:

$$_{15}P^{30} \rightarrow {}_{14}Si^{30} + {}_1e^0$$

The half-life period of the radioactive phosphorus is about 3 minutes.

Many other substances also can be made temporarily radioactive by bombardment with neutrons. In every case an unstable isotope is formed, and it explodes with the emission of a negative or positive electron, producing at the same time a stable nucleus.

555. Positrons.—A few months after the discovery of the neutron in 1932, a fourth fundamental particle was discovered by Anderson of the California Institute of Technology. He was investigating cosmic rays (§557) by means of a Wilson cloud chamber, placed in a strong, uniform magnetic field. Tracks were found that could only be explained by the presence of an electrically charged particle of the same mass as an electron, but with a positive charge equal in magnitude to that of an electron. It was therefore called a positive electron or, more briefly, a *positron.* Further study showed that similar particles could be produced by the impact of γ-rays on metals.

Figure 470 is from one of Anderson's photographs. At the upper right-hand corner a γ-ray photon, which does not itself produce a track in the cloud chamber, impinges on a plate of lead and produces a positron and an electron. The positron curves to the right in the magnetic field and the electron to the left. The latter passes through a second lead plate and thereby loses some velocity, so that its path in the magnetic field becomes more curved. (The change of curvature does not show clearly in this picture, because the electron was not moving in a plane perpendicular to the axis of the camera, but it was evident in other cases).

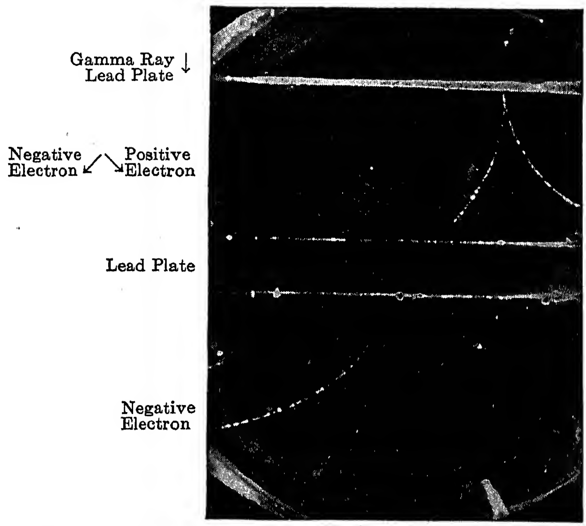


FIG. 470.—The production of an electron-positron pair by γ rays.

The production of an electron pair from a γ-ray photon is an example of the transformation of electromagnetic energy into mass, in accordance with the principle of relativity, which regards mass and energy as different aspects of the same entity. The energy of the γ-ray photon of frequency ν is $h\nu$ and its mass is $h\nu/c^2$, where c is the velocity of light in a vacuum and h is Planck's constant (§729). This, it is found, exactly accounts for the mass and energy of the positron and the electron.

The life of a positron is exceedingly short, of the order of a millionth of a second; for it quickly combines with an electron to produce photons of equivalent energy and mass, thus reversing the

process represented in the figure. This reverse transformation has also been found in cloud chamber photographs.

556. Structure of Nuclei.—When the negative electron and the proton were the only two fundamental particles known, it was generally believed that the nucleus of an atom is a group of protons and electrons, for example that an oxygen nucleus consists of 16 protons and 8 electrons. The weight of evidence now favors the belief that protons and neutrons are the fundamental entities from which nuclei are built. The atomic weights of most atoms are about twice their atomic numbers. This would signify that there are about an equal number of protons and neutrons in most nuclei.

As an illustration of the neutron-proton hypothesis, consider the structure of the ${}_3Li^7$ nucleus. The mass of a proton (on the basis of 16 as the atomic weight of oxygen) is 1.00807 and that of a neutron 1.00845, and the atomic weight of the lithium is 7.0164. If the nucleus consists of four neutrons and three protons its total mass should be 7.05801 which is greater than 7.0164 by 0.0416. This is the mass defect of the lithium nucleus. It is taken as the mass equivalent of the energy with which the component parts of the nucleus are bound together. Now the actual mass of the oxygen atom is 2.64×10^{-22} gram and the 0.0416 mass unit is found in grams by dividing by 16 and multiplying by 2.64×10^{-22}, which gives 0.69×10^{-24} gram. To reduce this to ergs we must multiply it by c^2 or 9×10^{20} and the result is 6.2×10^{-4} ergs. This is the amount of energy that must be supplied to the nucleus to separate it into its component parts. It is the amount of energy acquired by an electron in a potential fall of about 390 million volts. Thus the lithium nucleus is in no danger of spontaneous disintegration. Similar considerations apply to the nuclei of all stable elements.

557. Cosmic Rays.—The history of cosmic rays makes an interesting story. Discoveries in Physics are usually made in laboratories, but the study of cosmic rays has taken physicists with their instruments high into the air and deep into the earth and to various latitudes. The search began about 1900, when attention was directed in Germany and England to the question why the charge on an electroscope always leaks away, no matter what precautions or shields are used. At first many thought it was due altogether to radioactive emissions from the earth or the walls, but this was disproved when Gockel and Hess took electroscopes up in balloons and found that the leak actually increased with height.

It was then recognized that the origin was outside the earth, and the rays were christened *cosmic rays.*

It was found by Millikan and others that cosmic rays could pass through several hundred feet of water and corresponding thicknesses of denser substances, and from this the inference was drawn that they are of the nature of γ-rays but of far greater frequency. This was disproved by Compton and others by an extensive survey of the intensities of the rays at different latitudes, for it was found that the intensity decreases from the magnetic north pole to the magnetic equator. This effect of the earth's magnetic field is somewhat similar to what occurs in the case of the aurora borealis or "Northern lights," which is known to be due to the deflection of electrons from the sun by the earth's magnetic field and their concentration in spirals about the magnetic pole.

The weight of evidence at present is that cosmic rays, as they enter the atmosphere, consist (according to Compton) of positive and negative electrons with also a possibility of protons. These primary cosmic rays cause ionization in the atmosphere and produce also photons or very penetrating electromagnetic radiation. The separation of these secondary radiations from the primary cosmic rays is a very difficult problem and this accounts for the present lack of complete agreement on the subject. The analysis indicates that the primary cosmic rays have very high energies, of the order 10^{10} electron-volts (1 electron-volt being the energy an electron acquires in falling through a potential difference of 1 volt).

There is also a small directional asymmetry of cosmic rays ascribable to the motion of the earth with the rotation of the galaxy to which our sun belongs. These results agree with calculations made by Compton on the assumption that the primary cosmic rays are charged particles.

PROBLEMS

1. Two small spheres, each of 0.1 gm. mass, having equal charges, are suspended from the same point by silk fibers 80 cm. long. If the spheres are kept 8 cm. apart by repulsion, what is the charge on each? *Ans.* 17.7 e.s.u.

Electrostatic Fields.

2. Two charges +90 and −40 are 30 cm. apart. Find the intensity of field at a point in the line joining them 60 cm. from the negative and 90 cm. from the positive charge, and calculate the force on a charge of +20 if placed at this point.

3. Two small charged spheres repel each other with a force of 10 dynes when 2 cm. apart. If the charge on one of the spheres is doubled, and the distance between the spheres is doubled, what is the repulsion? *Ans.* 5 dynes.
4. What work is done in carrying a charge of 10 units from a point where the potential is 25 to a point where it is 40?
5. Find the electric field strength and potential at the corner C of a right triangle ABC if there are charges of +128 e.s.u. at A and −72 e.s.u. at B, where AC is 80 cm. and BC is 60 cm. *Ans.* 0.4 e.s.u.; 0.028 e.s.u.

Capacity. 6. Given two spheres of radii 3 cm. and 8 cm.; how will a charge of 66 units distribute itself over them if they are connected by a fine wire?

7. What is the charge on a spherical drop of water 2 mm. in diameter where the electric potential is 10? Two such charged drops unite to form a single spherical drop; assuming no charge is lost, what is the potential of the resulting drop? If three drops thus unite, what is the final potential? *Ans.* 1 e.s.u. of q; 1.6 e.s.u. of potential; 2.1 e.s.u. of potential.
8. A Leyden jar $\frac{1}{4}$ cm. thick is 3 cm. in radius and 9 cm. high. Find its capacity, if the dielectric constant for glass is 6. Find charge on each plate when *p. d.* is 15 e.s.u.
9. A condenser of 10 plates, each 20 cm. × 30 cm. has 0.4 mm. of air between each pair of plates. Find the capacity. *Ans.* 10,740 e.s.u.
10. Two plate condensers are joined in parallel. One is a 15 plate air condenser, each plate 11 cm. long and 5 cm. broad, 3 mm. apart; the other a mica condenser of 10 plates, 22 cm. long, 15 cm. broad, 0.5 mm. apart, specific inductive capacity of mica being 8. Find the capacity of each and the combined capacity. *Ans.* 204; 37,815; 38,019 e.s.u.
11. Two concentric spheres of radii 10 cm. and 10.3 cm. are separated by air and are charged to a difference of potential of 50 volts. Find the charge. *Ans.* 57.2 e.s.u.
12. A pair of circular plates of radii 10 cm. each are 2 mm. apart in air and are charged to a difference of potential of 20; when they are connected to the plates of an uncharged condenser, the difference of potential falls to 3. Find the capacity of this condenser. *Ans.* 708 e.s.u.
13. Find the capacity of a plate condenser made of two rectangular conductors 32 cm. long and 13 cm. broad, 0.2 cm. apart in air. *Ans.* 280 e.s.u.
14. If the air in Prob. 13 be replaced by 0.2 cm. sheet of glass of dielectric constant 7, find the charge on each plate when the difference of potential is 20 e.s.u. *Ans.* 39,200 e.s.u.
15. Find the work in ergs required to charge an isolated metal ball of 5 cm. radius with 20 e.s.u. of electricity. *Ans.* 40 ergs.
16. Find the intensity of the field at a point 40 cm. from a magnet in the perpendicular bisector of the line joining the poles of the magnet 6 cm. long and of pole strength 160 e.m.u. Calculate the force on a pole of +80 e.m.u. if placed at the point. *Ans.* .0148 oersted; 1.19 dynes.

Magnetic Fields.

17. A magnet NS 30 cm. long is held vertically and each pole has a strength of 9 units. What is the force on a unit pole at a point 20 cm. horizontally from the upper pole? What is the horizontal component of this force? *Ans.* .02 dynes; .019 dynes.

18. The point P is on the perpendicular bisector of a magnet NS, at a distance of 30 cm. from NS. A pole of strength 8 at P is acted on with a force of 3 dynes. Find the moment of the magnet NS. *Ans.* 10,125 e.m.u.

19. To hold a magnetic needle NS at an angle of 60° with the earth's field requires a torque of 0.6 dynes $\times$ 2 cm.; the horizontal intensity of the earth's field is 0.2. What is the moment of the magnet? *Ans.* 6.9 e.m.u.

20. A short bar magnet is placed with its axis perpendicular to the magnetic meridian, and with the line of the axis passing through the center of a compass needle. At a station X, the compass needle is deflected through an angle ϕ, when the center of the magnet is 40 cm. from the center of the needle. At a station Y, the distance from magnet to needle is 35 cm. for the same deflection ϕ. Compare the horizontal intensities.

Ans. Ratio 1.49.

21. The horizontal intensity of the earth's field at Indianapolis is 0.2, and at Minneapolis it is 0.18; if a magnetic needle makes 100 vibrations in five minutes at Indianapolis, what will its period be at Minneapolis?

Ans. 3.16 sec.

22. To deflect a suspended magnet through an angle of 20° from the magnetic meridian requires 180° of torsion in the wire suspension; how many degrees of torsion must be given the suspension to produce a deflection of 45° from the magnetic meridian?

23. A horizontal magnetic needle makes 40 oscillations per minute at a place where the dip is 70°, and 50 oscillations per minute where the dip is 60°. The total intensity at the first place is 0.6; what is it at the second place?

Ans. 0.64 oersted.

24. The center of a short bar magnet is at the corner A of a square $ABCD$ and its axis is in line with the side AB. The moment of the magnet is 400, and the length of one side of the square is 60 cm.; find the intensity of the magnetic field at the corners B and D, due to the bar magnet.

Ans. .0037; .00185 oersted.

25. A circular coil of 30 cm. diameter has 20 turns. Compute the intensity of the magnetic field at the center when a current of 10 amperes flows through the coil. *Ans.* 8.4 oersteds.

Magnetic Fields of Currents.

26. Find the field strength 16 cm. from the center of a coil of twenty turns in the line of its axis if the coil carries 0.5 amp. and is 24 cm. in diameter.

Ans. 0.113 oersted.

27. Find force on a pole of 30 e.m.u. if placed at center of coil in problem 26.

28. A circular coil was placed at right angles to the magnetic meridian. The number of oscillations of a small magnetic needle at the center was counted: (a) when there was no current in the coil; (b) when a current i_1 was sent through the coil; (c) when a current i_2 was used. For (a) there were 40 oscillations per minute, for (b) 30 oscillations per minute, and for (c) 20 oscillations per minute. What was the relative strength of the currents i_1 and i_2?

29. A slender solenoid has a length of 50 cm. and has 300 turns of wire. What is the field at the center when the current in the coil is 7 amperes?

Ans. 52.8 oersteds.

30. Calculate the current which produces a magnetic field in the middle of a slender solenoid equal to the earth's field of 0.6, the solenoid being 80 cm. long, and having 400 turns.

31. A toroidal solenoid ring has 20 cm. mean diameter, and a cross-section of 18 sq. cm. The coil has 600 windings and carries 10 amperes. How great is the magnetic induction B? How many magnetic lines are produced? (The permeability $\mu = 200$.) *Ans.* $B = 24{,}000$; $N = 432{,}000$.

32. Calculate current which will deflect a tangent galvanometer 45°, if the galvanometer consists of a coil 18 cm. in diameter, of 7 turns of wire, set up in a field of 0.198 lines per cm². *Ans.* 0.4 amperes.

Galvanometers.

33. The coil of a tangent galvanometer is 34 cm. in diameter with one turn and carries a current of 15 amperes. What is the torque on a needle of moment 1.5 at the center?

34. A coil of a tangent galvanometer is to have 10 turns; what should the radius of the coil be, so that the tangent of the angle of deflection of the needle will give the current directly in amperes, at a station where the intensity of the earth's field is 0.19? *Ans.* 33 cm.

35. A current of 6 amperes flows for 4 min. in a circuit of 12 ohms resistance; what is the e.m.f. required? What is the total work done in ergs, also in joules? What is the power in watts?

Work and Ohm's Law.

36. Three electromagnets of resistances 50, 76 and 11 ohms respectively are joined in multiple arc, and a total current of 2 amperes flows through the three; what is the current in each? *Ans.* .32; .21; 1.47 amperes.

37. A wire 0.3 mm. in diameter and 3 meters long has a resistance of 4.48 ohms at room temperature (20°C.) and of 6.28 ohms at 95°C. Calculate (a) the resistance at 0°C., (b) the specific resistance of 0°C., and (c) the temperature coefficient.

Ans. 4.00 ohm; 9.4×10^{-6} ohm cm.; 0.006 per degree C.

38. A multiple arc has branches of 50, 30 and 10 ohms. A fourth branch is put in so that the total resistance is 2 ohms. What is the resistance of the fourth branch? *Ans.* 2.88 ohms.

39. A generator delivers 100 amperes at 110 volts; what is the power in kilowatts and what in H. P.?

40. The resistance of a galvanometer is 126 ohms, and a shunt of 14 ohms is put in; what is the resistance of the shunted galvanometer?

41. By experimenting with a Weston ammeter it was found that 0.00013 amperes through the coil gave one unit scale-deflection. If the resistance of the coil circuit be 5.60 ohms, what must be that of the shunt so that .01 ampere in the external circuit will give 1 unit scale deflection?

Ans. .0738 ohm.

42. It is desired to supply 600 incandescent lamps, in parallel, with $\frac{1}{2}$ amp. each, at 110 volts potential difference between the lamp terminals. If the drop in the line be 2.2 volts, what is the resistance of the line and how much power is lost in it? How much power must be generated and what voltage? *Ans.* .0073 ohms; 660 watts; 33.66 kw.; 112.2 volts.

43. A car is lighted by five lamps of 220 ohms resistance each, joined in series. What is the total resistance of the lamps? If the difference of potential

between the ends of the lamp circuit be 550 volts, what current flows through the lamps? What power is expended in this circuit and at 9 c. per kilowatt hr., what does it cost to light a car for one hour?

Ans. 1100 ohms; 0.5 ampere; 275 watts; 2.48 cts. per hr.

44. If the motive circuit of a snow sweeper takes 50 amp. (at 550 volts) and the broom motors takes 80 amp., find the total power consumed in the car if the two circuits be in parallel across 550 volt mains. Find cost per hour at 9 c per kilowatt hr. *Ans.* 71.5 kw. \$6.43.

45. Ten storage cells, each of 2 volts e.m.f. and 0.2 ohm resistance, are arranged in two parallel lines, each with five cells in series. The terminals of this battery are connected by two lead wires, each of 0.7 ohm resistance, to a multiple arc, consisting of three coils of resistance 3, 4, 24 ohms in parallel. Calculate (*a*) the resistance of the arc, (*b*) the resistance of the battery, (*c*) the total resistance of the circuit, (*d*) the current in the circuit, (*e*) the current in each lead wire, (*f*) the current in each cell, (*g*) the current in each coil, and (h) the p.d. at the terminals of the battery.

46. An 8 volt battery of 4 ohms resistance is put in parallel with a 6 volt battery 2 ohms resistance (like poles joined) and the terminals are connected to the ends of a 2 ohm resistance. Calculate the current in each battery and the current in the resistance. *Ans.* 1; 1; 2 amp.

47. Find the resistance of a tube of mercury at 0°C., 1 meter long, and 1 cm. in diameter. *Ans.* 0.012 ohm.

48. The resistance of a certain firm's copper wire 1 foot long and a mil (one thousandth of an inch) in diameter is 10.7 ohms. What is its specific resistance in ohms per cm./cm.^{2}? *Ans.* 1.78×10^{-6} ohms per cm./cm.^{2}

49. The specific resistance of copper is 1.5×10^{-6} and of aluminum is 3.2×10^{-6}. With copper at 18 c. per pound, what must be the price of aluminum to compete as an electrical conductor? *Ans.* 27.6 cts. per lb.

50. Given 3 cells of 1.4 volts and 0.8 ohms resistance each, find resistance of the battery if the cells be connected in series and calculate the current through an external resistance of 9 ohms. *Ans.* 2.4 ohms; 0.368 amperes.

51. Find the resistance of the battery above if cells be in parallel and also the current when the external resistance is 9 ohms.

Ans. 0.27 ohms; 0.15 amperes.

52. Given 20 cells, each with an e.m.f. of 1.7 volts, and an internal resistance of 3 ohms. Calculate the current in the following cases:

(*a*) External resistance $R = 100$ ohms, cells in series.

(*b*) $R = 100$ ohms, cells in parallel.

(*c*) $R = 20$ ohms, cells in series; also in parallel.

(*d*) $R = 20$ ohms, 4 parallel rows of 5 cells in series.

(*e*) Arrangement for maximum current through 20 ohms.

53. A milli-ammeter, which is to be used as a voltmeter, indicates .005 ampere for a scale division, and has a resistance of 40 ohms. There are 50 scale divisions. What resistance in series with the instrument will enable it to be used for measurements up to 300 volts? *Ans.* 1160 ohms.

54. It is required to generate 1000 calories of heat per minute in a coil, the e.m.f. at the terminals of the coil being 110 volts; what resistance must the coil have? *Ans.* 173.4 ohms.

Joule's Law. **55.** The same current flows through a platinum wire 25 cm. long and 0.5 mm. diameter, and through a copper wire 500 cm. long and 0.6 mm. diameter. What are the relative heat quantities developed in these wires? (Sp. res. of Cu = 1.5×10^{-6}; sp. res. of Pt = 8.9×10^{-6}). *Ans.* 0.43.

56. A uniform current flows for 10 minutes and deposits 4 grams of silver; calculate the current. *Ans.* 5.9 amperes.

57. How much copper can a dynamo giving 30 amperes deposit in an hour?

Electrolysis. **58.** How many cubic centimeters of hydrogen at 0°C. and 76 cm. Hg. pressure can a current of 30 amperes produce by the decomposition of acidulated water in an hour? (Sp. gr. of H_2 = .000089 at 0°C. and 76 cm. Hg. pressure.) *Ans.* 12,700 c.c.

59. Show that Lenz's law and the rule for the direction of an induced current in a conductor moved across a magnetic field agree.

Electromagnetic Induction. **60.** The diameter of a circular ring is 30 cm. and the resistance is 0.1 ohm. Find the quantity of electricity in coulombs which will flow in the ring when revolved from a position at right angles to a magnetic field to a position parallel to the field. $H = 20$. *Ans.* .0014137 coulombs.

61. A circular coil 40 cm. in diameter and with 100 turns is rotated five times per second about a vertical diameter as axis. Find the maximum e.m.f. induced. The horizontal component of the field is 0.2. *Ans.* 0.0079 volt.

62. If the angle of dip is 70°, what is the maximum e.m.f. induced when the coil above is rotated about a horizontal axis parallel to the horizontal component of the field ten times per second? *Ans.* 0.0432 volt.

63. Calculate the e.m.f. induced in a car axle 120 cm. long having a horizontal linear velocity of 25 meters per second, where the total intensity of the field is 0.6 and the angle of dip is 70°. *Ans.* 0.0169 volts.

64. A solenoid has a self-inductance of 10 millihenries and carries a current of 4 amperes. What is the energy of the field?

65. A solenoid, of 1200 turns wound on a wooden rod 80 cm. long and 5 cm. in diameter, has a secondary coil of 100 turns and 0.5 ohm resistance wound around its center. What quantity of electricity flows in the secondary when a current of 3 amperes in the primary is reversed? *Ans.* 0.0178 coulomb.

66. Draw a figure showing the directions of the induced currents in the disk of a pendulum swinging between the poles of a magnet across the field. How should the disk be laminated to make the induced currents a minimum?

67. A rectangular coil, 10 cm. $\times$ 12 cm., can rotate about a vertical axis which bisects the 12 cm. sides. A current of 3 amperes flows through the coil, and the horizontal intensity of the magnetic field is 0.2. What is the torque (moment of force), when the plane of the coil is parallel to the field? *Ans.* 7.2 dyne-cm.

General Problems. **68.** Two bar magnets, bound together but in opposition, when suspended have a period of 10 seconds, but when the weaker one is reversed in the system the period is 2 seconds. If both have the same dimensions find the ratio of their magnetic moments.

69. If it requires a torque of 20 gm. cms. to hold a suspended magnet at an angle of 60° from the direction of a magnetic field of intensity 5 units, what is the magnetic moment of the magnet?

70. A rectangular bar magnet having a length of 10 cm., width 2 cm., thickness 2 cm., and mass of 240 gms. makes 20 vibrations per minute in a magnetic field of strength .25 oersted. What is its pole strength?

71. A tangent galvanometer has two sets of coils. A current through the first coil produces a deflection of 20°, and the same current through the second coil produces a deflection of 40°; the first coil has 10 turns; how many turns are in the second coil?

72. What resistance would be required in series with a galvanometer which has a resistance of 200 ohms and a shunt of 10 ohms, in a circuit with an electromotive force of 2 volts, in order that the current through the galvanometer may be 10^{-5} ampere?

73. Two concentric loops of wire of 10 and 20 cm. radius respectively are connected in series. What field strength do they produce at the center if a current of 2 amperes flows through them: (*a*) in the same direction; (*b*) in opposite directions?

74. An electric circuit made up of a cell of internal resistance 1 ohm in series with two wires of 2 ohms and 3 ohms respectively, carries a current of 0.3 ampere. What is the e.m.f. of the cell? What is the potential difference at the terminals?

75. When a cell is short circuited by a resistance of 10 ohms, the current is 0.1 amp.; when short circuited by a resistance of 25 ohms the current is 0.05 amp. What is the resistance of the cell?

76. A copper telephone line (two wires) has a resistance of 3.6 ohms when measured at the station. Where is the line short-circuited, if the wire is of cross-section 0.02 sq. cm. and specific resistance 1.5×10^{-6} ohm-cm.?

77. A "one ohm" coil is tested and found to have a resistance of 1.004 ohms. The resistance of No. 30 constantan wire is 9 ohms per meter. What length of the latter must be used in parallel with the coil to correct it?

78. A copper plating cell is placed in parallel with a nickel plating cell of equal electrical resistance. In series with the pair is a cell containing a silver solution. How much Cu and Ni will be deposited when 5 gm. of Ag are deposited? What should be the ratio of the resistances of the Cu and Ni cells in order to make the mass per second deposited the same for each?

79. A toroidal solenoid has an average circumference of 40 cm. and cross section 2 cm. in diameter. The coil wound on the ring has 150 turns, and the permeability of the iron core is 900. Find the current to produce a flux of 10,000 lines across any section.

SOUND

By A. Wilmer Duff, D. Sc.

Professor Emeritus of Physics in the Worcester Polytechnic Institute

NATURE AND PROPAGATION OF SOUND

558. Sources of Sound.—On hearing a sound, we instinctively think of its origin, and we are usually able to trace it to some body, which we call the source of the sound. To ascertain how a body produces a sound, we may take a body that can be kept sounding while it is being observed. If a large bell be made to produce sound by striking it, or a glass jar by stroking it with a violin bow, a light pendulum hung against the bell or jar will be kicked away at each contact. A metal rod, clamped at the middle and sounded by rubbing one-half of it with a rosined cloth or glove, will give violent blows to a pendulum, hung in contact with the other end. If a violin be held in the hand and one of the strings be plucked or stroked by a bow, the hand will tell us that the wooden body of the violin is vibrating, and while the vibrations of the body cannot be seen, those of the string are clearly visible.

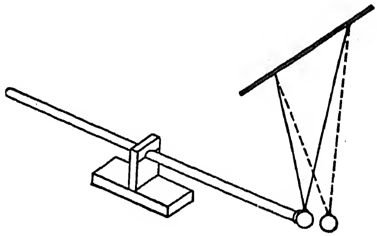


Fig. 471.

In all cases in which the action can be observed closely enough, sounding bodies are found to be in a state of *vibration*. In many cases the vibrations are so small or cease so quickly that they cannot be detected. But there is no doubt that the head of a hammer vibrates for a moment when it strikes a nail, and the board into which the nail is being driven also vibrates for a moment.

559. Media in Which Sound Travels.—It is well known that the clearness with which distant sounds are heard depends on the state of the atmosphere and the direction of the wind. Hence the air is the ordinary medium of transmission. Sound will not

travel in a vacuum, as can be readily shown by placing an electric bell under the receiver of an air pump. The sound will diminish as the air is removed, but it will be restored if the air or any other gas is allowed to enter.

Sound is also transmitted by liquids and solids. If two stones be struck together under water, a loud sound will be heard by an ear held beneath the surface. A watch placed on one end of a long table can be heard by an ear pressed against the other end of the table. Miners, imprisoned by an accident in a mine, sometimes send signals to the outer world by blows of their picks on the rock. The approach of a distant train or the galloping of

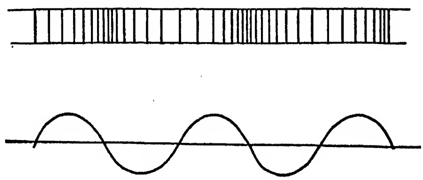


Fig. 472.

horses can be heard by an ear pressed against the ground. Beethoven, who was deaf, heard some of his own composistions only by means of a stick, one end of which rested against the sounding board of the piano, while the other was pressed against his teeth.

560. The Nature of Sound.—Since the sources of sound are vibrating bodies, sound travelling through a medium must be the transmission of vibrations, that is, a wave motion. It cannot consist, like odors, of particles transmitted from the source, for it can travel through partitions.

Wave motions may be either *transverse*, like the to and fro motions of a cord, or *longitudinal*, like the compressions and extensions of a spiral spring, or they may be a mixture of both, as in the case of water waves. Now a gas offers elastic resistance to compression and can, therefore, transmit longitudinal waves; but it offers no sustained resistance to changes of shape, or shears, and therefore it cannot transmit transverse elastic waves. Hence sound waves are longitudinal waves, or waves of compression and dilatation.

To represent sound waves graphically (Fig. 472) we use the device explained in §238, that is, we take the axis of abscissæ in the direction in which the waves are travelling, and draw ordinates to represent the displacements of the air particles at the

corresponding points, although the displacements are really in the direction of propagation. These actual displacements are usually very minute, ranging from 2 or 3 mm. in the loudest sounds to about 10^{-7} mm. in sounds that are just audible. Hence the ordinates are drawn on a greatly enlarged scale. The forward motion of the wave is represented to the mind by supposing the curve to move forward with the velocity of the wave. The curve crosses the axis of abscissæ at the middle of a condensation or of a rarefaction, and the maximum ordinate occurs at a place of no condensation or rarefaction. Sound waves are sometimes represented by similar curves but with ordinates drawn to represent the degrees of condensation (positive or negative) at points in the wave. Such curves are a quarter of a wave length ahead of those that respresent displacements.

That sound has all the properties characteristic of wave motions in general will appear in the paragraphs which immediately follow.

561. Velocity of Sound.—When a man is seen chopping a log or hammering a nail at a great distance, the sound of each blow is not heard until some time after the blow is seen. Steam may be seen to issue from a distant whistle before the sound is heard. Fire-alarm whistles, sounded simultaneously, may be heard separately, if the observer is not at equal distances from the stations. Lightning precedes thunder. From such facts it is evident that sound travels at a definite rate that can be measured. That this rate is not as great as the velocity of shells fired by modern high power guns is shown by the experiences of soldiers in trenches. The sound made by a passing shell is often heard before that of the firing of the gun arrives.

The velocity of sound has been determined by methods suggested by some of the above experiences. The most accurate of the early direct determinations were made near Paris in 1822 and near Utrecht in 1823, and gave a mean value of about 341 meters per second at 16°C. The sounds observed were those of the firing of cannon at great distances, and care was taken to reduce the effect of wind by alternate observations of sounds travelling in opposite directions. The velocity of sound in water was found in 1827 by means of bells, sounded beneath the water of Lake Geneva in Switzerland. But we shall see later that there are laboratory methods more readily available for finding the velocity of sound in gases, liquids, or solids.

That **sounds of different pitch travel at practically the same speed** is shown by the fact that music by a band can be heard as music at a considerable distance; for, if notes sounded at the same time were not heard simultaneously, both harmony and melody would be distorted. There is evidence that, near the source, very loud sounds (explosive-waves) travel at more than the ordinary velocity.

From the principles of Mechanics, applied to waves of compressions and rarefactions, Newton derived the formula which has been stated in §251, namely,

$$v = \sqrt{\frac{E}{\rho}}$$

ρ being the density of the medium and E its elasticity. Now, the elasticity of a gas at constant temperature, or its *isothermal* elasticity, is equal to the pressure, p (§223). But if we substitute p, expressed in dynes per square centimeter, in the above, we get a result that is much too small, as Newton found. This difficulty was not removed, until Laplace pointed out that the temperature must be momentarily elevated in a compression and depressed in a rarefaction, as a train of sound waves passes, and that E should, therefore, be taken as the *adiabatic* elasticity, which, as we have seen (§346), is κp, where κ is the ratio of the specific heats of the gas at constant pressure and at constant volume, respectively. With this correction, the formula becomes

$$v = \sqrt{\kappa\frac{p}{\rho}}$$

and this is found to agree with experimental results.

From this formula it is evident that **the velocity of sound in a gas is independent of the pressure or density, provided the temperature is constant,** since, in accordance with Boyle's law, p is proportional to ρ when the temperature is constant. Thus, at the same temperature, the velocity is the same at the top of a mountain as in a valley, if the constitution of the air is the same. In gases of different densities the velocities are inversely as the square roots of the densities. Thus the velocity in hydrogen is about four times that in oxygen. Since the density of water vapor is less than that of air, the presence of water vapor in air, at given

atmospheric pressure, causes a slight increase in the velocity of sound.

From the formula we can also find how the velocity of sound depends on the temperature of a gas. In the general gas law (§280) let V be the volume of unit mass. Then $V = 1/\rho$, where ρ is the density, and $(p/\rho)_t = (p/\rho)_o\,(1 + at)$. Hence

$$v = \sqrt{\kappa\left(\frac{p}{\rho}\right)_o(1 + at)} = v_o\sqrt{(1 + at)}$$

where $v_o = \sqrt{\kappa\left(\dfrac{p}{\rho}\right)_o}$ is the velocity at 0°C. and is independent of the pressure.

VELOCITY OF SOUND IN DIFFERENT SUBSTANCES (IN METERS PER SECOND) AT 0°C.

Air	332	Steel	4975
Hydrogen	1268	Lead	1420
Carbon dioxide	261	Glass	4860
Fresh water	1435	Pine wood	3300
Sea water	1454	Walnut wood	4800
Mercury	1484	India rubber	5000

562. Mechanical Effects of Sound Waves.—Sound waves produce mechanical effects by which they can be detected and, to some

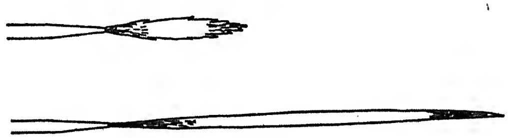


FIG. 473.

extent, measured. A very simple and useful detector is a "sensitive flame," that is, a long slender gas jet issuing from a very fine nozzle (a glass tube drawn out to a fine point) under steady pressure. If the flow is regulated until the flame is just about to flare or become unsteady, sounds of high pitch, falling on the orifice, will cause instability, and the flame will shorten greatly and "roar," like an ordinary gas flame when the stopcock is too wide open. A hiss or the rattling of a bunch of keys is especially effective (Fig. 473).

A so-called "manometric flame" is a gas jet, fed by gas which passes through a small box, one side of which is of thin rubber. Sound waves, falling on the rubber, produce variations of pressure

in the gas and corresponding variations in the height of the flame. If the latter be viewed in a rotating mirror, a band of curved outline corresponding to the peculiarities of the sound will be seen (Fig. 474).

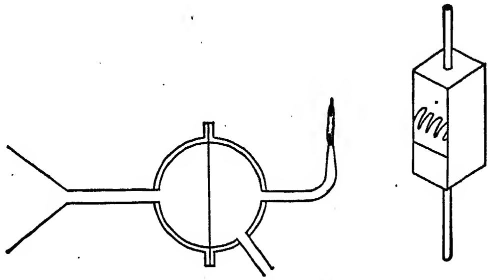


FIG. 474.

The vibrations of a disk on which sound waves fall can be made visible by reflecting a beam of light from a small mirror connected with the disk. This method has been very successfully used, by Professor D. C. Miller in his "phonodeik." A fine fiber f, attached at one end to a thin glass plate p, is wrapped around a tiny spindle, which carries the mirror m; a record obtained by Miller will be given later (Fig. 484).

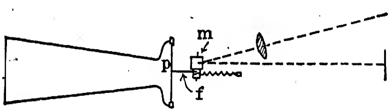


FIG. 475.

The effect of sound waves on the disk of disk of a telephone consists of vibrations of the disk, produced by the alternating pressure of the waves. The vibrations of the disk produce alternating electric currents in the line (§503), and these can be shown by a vibration galvanometer (a galvanometer with a very light vibrating system that can follow the changes in an alternating current), or they can be rectified by a crystal rectifier (§525) and shown by an ordinary galvanometer.

Sustained waves of any kind carry a steady stream of energy from the source and, falling on a surface, produce a sustained pressure. In the case of sound waves this pressure is small. It can be shown by concentrating the loud sound produced by a stream of sparks from an induction coil, by means of a large metallic mirror. A small vane, similar to that of a radiometer, placed at the focus, will be set in rotation. Waves falling on a small disk in a resonating tube tend to set it at right angles to the stream, by an action somewhat similar to that described in §204. This action has been used, in certain investigations on the energy flow in sound waves, by Lord Rayleigh and others, a beam of light being reflected from the disk to show the amount of deflection.

563. Location of Guns by Sound.—"Sound-ranging," or the determination of the position of a gun from the times of arrival of the sound at three or more stations, played an important part in the war of 1914–1918. Let A, B, C (Fig. 476) be three stations at which the arrival of the sound is recorded photographically by some instrument (such as the one described in §562), and let O be the position of the gun. The difference of the times of arrival at A and B gives us $(OA - OB)$. Now a curve on which $(OA - OB)$ is constant is a hyperbola, with A and B as foci. Similarly from the times of arrival at B and C we get a second hyperbola, and O is at the intersection of the two curves. In practice more rapid methods than the drawing of the curves were used, and allowance was made for wind and temperature. In this way a gun could be located to about 50 ft. in 6 miles, or about as accurately as a gun could be aimed to strike a target placed at an equal distance.

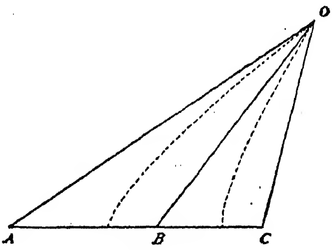


FIG. 476.

564. Photographs of Sound Waves.—Sound waves can be photographed by a method very analogous to the method used in projecting ripples (§256). The waves pass in front of a photographic plate in the dark; a single spark from a distant Leyden jar illuminates the plate for a moment, and leaves a record of

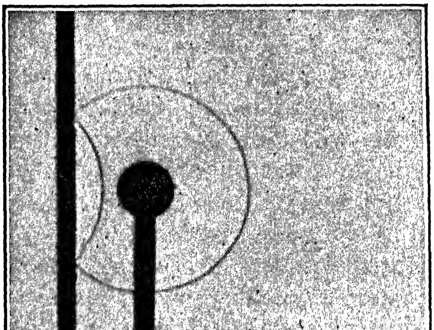


FIG. 477.

the condensations and rarefactions, owing to the refraction of the light as it passes through the sound wave. Figure 477 is such a record (by Foley and Souder), showing the reflection of a spherical wave of sound by a plane surface. (See also §667.)

565. Reflection of Sound.—Sound, like other wave motions, is reflected when it falls on a suitable surface, and it then follows the

same laws as light and radiant heat, namely, the reflected and incident rays lie on opposite sides of the normal to the surface and make equal angles with it. The identity of these laws for sound and light can be shown by means of two large concave metallic mirrors, placed opposite each other, with their axes in the same line. In front of each mirror there is a point, its principal focus, such that a bright light placed at one focus produces a bright image at the other. When these points have been found, a source of sound of high pitch placed at one focus will produce an intense sound at the other focus, as may be seen by its action on a sensitive flame. If a singing arc light be placed at one focus, the light, heat, and sound will

FIG. 478.

be focussed simultaneously by the other mirror, as may be ascertained by means of a small screen, a thermopile (§332), and a sensitive flame respectively.

Reflection of sound gives rise to echoes. The echo from a large reflecting surface, such as the side of an isolated house, is heard most distinctly when the observer is in a line perpendicular to the surface, and it decreases rapidly as he moves away from that line. Most of the sound heard in an auditorium is reflected sound. This will be referred to more fully later (§596).

566. Refraction of Sound.—Waves are refracted, or their line of propagation is bent, when they pass obliquely from one medium into another in which the velocity is different (§255). The refraction of sound cannot be shown satisfactorily on a small scale by lecture or laboratory apparatus, but it takes place on a large scale in nature. The chief causes of such refraction are winds and variations of temperature in the atmosphere.

The velocity of a wind is usually less nearer the surface of the earth than higher up, since, near the surface, it is retarded by the frictional resistance of the surface. When sound is travelling in the same direction as the wind, its resultant velocity is greater above

than below. Hence the waves, which always travel at right angles to their fronts, are tilted forward, or the direction of their motion is deflected downward (Fig. 479, *B*). The opposite effect takes place when the sound is travelling against the wind (Fig. 479, *A*). This explains why sound is better heard with the wind than against it, and why it is an advantage in the latter case to listen from an elevation.

Similar effects result from the temperature being different at different heights in the atmosphere. Usually on fine days the temperature is lower at an elevation, and the velocity of sound is less there. Thus the waves are deflected upward, as in Fig. 479, *A*, and hearing near the surface is poor. At night or near sunrise or sunset, the surface is colder, and the gradient of temperature is less and sounds are heard better. Moreover, on a hot day the

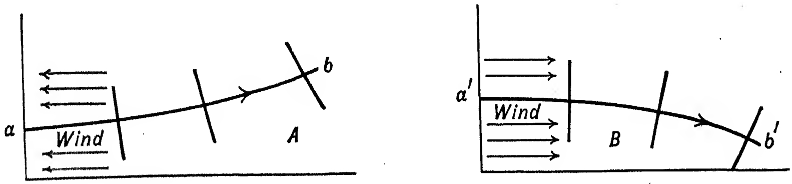


FIG. 479.

air is heated irregularly by contact with parts of the surface at different temperatures, and there are numerous irregular convection currents in the atmosphere. These break up the waves by causing irregular reflections and refractions of sound, as in the similar case of light passing over a heated stove. Distant sounds are frequently heard most clearly before a storm or when the ground is covered with snow, because the air is more uniform and undisturbed.

Loud sounds, such as those from a steam siren or a large gun, do not die away regularly with distance along the earth's surface. For example in the bombardment of Antwerp in 1914 there was a zone of silence from about 35 to 60 km. Such abnormal effects seem to be due to a reversal of the temperature gradient in the atmosphere, so that at a height of 20 or 30 km. the temperature may be greater than at intervening heights and approximately the same as at the surface of the earth. Sound waves ascending obliquely to such a height would gradually curve around by refraction and return to the earth's surface beyond a soundless zone. The study of

meteors has also indicated the existence of such a reversal of temperature gradient, and a sufficient cause is found in the absorption of sunlight by an ozone layer at a still greater height (§532).

That there is considerable damping, even in still air and especially for sounds of high pitch, was discovered in 1898 by the author, who found that the damping was very much greater than could be accounted for by viscosity and heat conduction between condensations and rarefactions, two reasons for expecting damping that had been treated by Stokes and Kirchhoff. As an explanation Lord Rayleigh suggested in 1900 that damping would result from the time required for the translational kinetic energy of molecules to be changed in part, by impacts, into rotational energy. Damping of about the same order was again found by Knudsen in 1931 by the study of the decay of reverberation in a closed room, and Rayleigh's suggestion has been developed by others. Damping is much less for sounds of low pitch, but for supersonic (inaudible) sounds of extremely high pitch Pierce found it so great as to make some gases nearly opaque to such sounds.

567. Diffraction of Sound.—Diffraction means the bending of waves around obstacles (§256). It prevents the formation of sharp shadows. The amount of diffraction in any case depends on the linear dimensions of the obstacle compared with the wave length of the sound. A hill casts a fairly definite sound shadow, because it is large in comparison with the wave length. Cases have been known in which houses in the shadow of a hill have suffered no damage from very loud sounds, such as the explosion of a powder magazine or the firing of cannon, while the windows of equally distant houses, not in the shadow, were broken by the impact of the sound waves. But small obstacles, such as trees and posts, cast no observable sound shadows, except when tested by waves of very short length. The human head is of sufficient size to cast something of a shadow for sounds of short wave length. If the sound is to the left or the right of the observer, one ear is in this partial shadow, and, from the difference of the intensities, as heard by the two ears, we judge of the direction of the source. Long waves produce about equal intensities at the two ears, but at one ear the phase of the waves is later than at the other, and it has been shown that it is by this difference of phase the mind unconsciously judges of the direction of the source of the sound. For sounds of intermediate wave length both indications are used, and in any case the final determination of the direction of the source is usually made by turning the head until the sound is directly ahead.

568. The Phonograph and the Gramophone. Acoustic Impedance.—When we speak against the middle of a thin elastic disk, such as is used in the telephone transmitter, the disk vibrates backward and forward in unison with the sound waves. In the phonograph these vibrations are recorded on a drum of hard wax that is kept in rotation behind the vibrating disk. A short needle attached to the vibrating disk presses on the moving drum, and the vibrations cause it to plough a fine furrow in the wax. Thus the furrow is a record of the vibrations of the disk. The sound can be reproduced by allowing the needle to travel again along the furrow. It is thus pushed up and down, following very closely the motion that produced the furrow, so that it causes the disk to repeat its original vibrations, and the disk therefore reproduces the original sound. (It is found better to use different disks and needles for recording and reproducing.)

The principle of the gramophone is essentially the same, but the needle is connected to the center of the vibrating disk by a lever in such a way that it moves sidewise and not up and down. It thus produces a transverse furrow on a rotating disk.

The gramophone has been greatly improved by a study of its parts in the light of a striking parallelism between a train of acoustic mechanisms and the parts of an electrical network. In this relationship inertia corresponds to inductance (electromagnetic inertia), the reciprocal of rigidity (pliancy) to capacity, frictional resistance to electrical resistance, so that there is an *acoustic impedance* corresponding exactly to electrical impedance. In this way problems in acoustic design can be translated into electrical problems, the solutions of which are known or can be found by familiar methods. The basis of the parallelism is the existence of a fundamental differential equation that is the same in form for mechanics and acoustics.

MUSICAL SOUNDS

569. Characteristics of Musical Sounds.—The ear is remarkably acute in distinguishing minute differences in sounds. In doing so it takes note of three fundamental properties in which sounds differ, namely, *loudness*, *pitch*, and *quality*. These we may call the three characteristics of musical sounds. They and the direction from which the sound comes are all that the ear can tell us about it. From these the mind can, as a result of long experience, draw very rapid conclusions as to the nature of the source. The words loudness, pitch, and quality need no definition, as we are more familiar with them than with the terms which might be used in defining them. But we must remember that the words stand for sensations. Sound waves must have corresponding characteristics that account for these differences in the sensations of sound.

570. Loudness.—If we strike a bell, a drum or a violin string very gently, the vibrations of the instrument will be of small amplitude and the sound will be weak, but a stronger blow will produce vibrations of greater amplitude. Now it is evident that the amplitude of the air vibrations is greater the greater the amplitude of vibration of the source. Hence the *loudness* of the sound heard depends on the amplitude of the vibrations in the waves. The same conclusion is reached by considering that the sound heard is weaker the farther the hearer is from the source. The energy that falls on the ear must be less at greater distances, and therefore the amplitude of the vibrations must be less.

The *intensity* of sound waves means, as in the case of waves of any kind (§259), the flow of energy per unit time per unit area perpendicular to the direction in which the waves are travelling. Hence, in the case of strictly spherical waves spreading from a point source, the intensity would vary inversely as the square of the distance, provided the waves were not damped or scattered in some way. But the sound that comes directly from the source to the ear cannot usually be separated from that heard by reflection from neighboring surfaces, including the ground. Thus the inverse square law is usually not directly applicable to calculating the intensity of sound.

Waves of compression and rarefaction in air are not always audible as sound. If the intensity is too small, no sound is heard; if

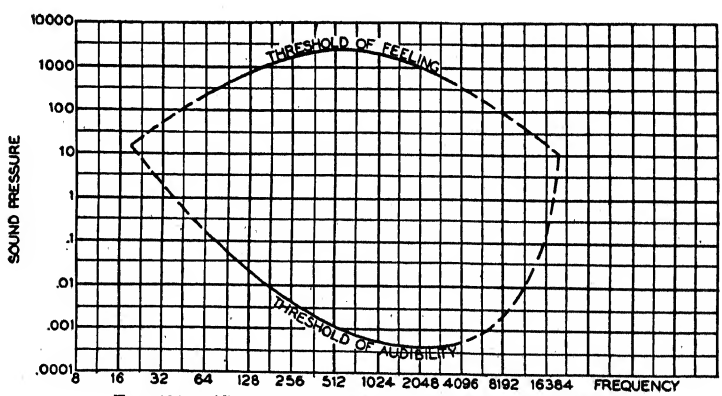


FIG. 480.—(Courtesy of the Bell Telephone Laboratories.)

it is too great, the effect ceases to be sound and becomes a feeling of pain in the ear. The limits, called the threshold of audibility

and the threshold of feeling, depend on the frequency and are different for different observers. Figure 480 shows the average results for a large number of observers, and it will readily be seen how the limits vary with the frequency. The numbers at the left are the averaged oscillatory pressures, in dynes per cm.^{2}, exerted by the waves on an obstacle or on the ear. A pressure of 1 dyne per cm.^{2} is common in ordinary conversation. It is about a millionth of an atmosphere, and the sensitiveness of the ear extends to pressures about $\frac{1}{500}$ of that.

Loudness is subjective, it refers to a feeling or sensation; intensity is objective; it is measured by instruments. A way of relating them has been found very useful by radio engineers and others. It had long been known that, if I is the intensity of a sound of unvarying pitch and quality, the smallest increment of it, says ΔI, that can be detected by the same observer under the same circumstances is proportional to I, so that $\Delta I/I$ is the same for different values of I. While this law is defective for very great or very small intensities, it suggests (by integration) that if I and I_0 are *any* two intensities of a sound of given pitch and quality, their difference in loudness may be taken as measured by $\log_{10} (I/I_0)$. This is now the practice, and the unit of loudness-difference is called the *bel*, in honor of Graham Bell, the inventor of the telephone. However, the *decibel*, one-tenth of a bel, is found more convenient as a unit. Thus, if the loudness-difference is α decibels,

$$\alpha = 10 \log_{10} (I/I_0)$$

If $I = 2I_0$, $\alpha = 3.0$ decibels; and if $\alpha = 1$ decibel, $I = 1.26\, I_0$.

To make the decibel scale definite, an intensity of 1 microwatt (10^{-6} watt) per cm.^{2} is usually taken as the standard. If we now take I_0 in the equation as 1 microwatt per cm.^{2} and measure I in microwatts per cm.^{2},

$$\alpha = 10 \log_{10} I$$

and α now means decibels of loudness above or below the standard, taken as zero of the scale (like temperature above or below zero, or land elevation above or below sea-level).

In defining loudness-difference in this way, we are giving the word "loudness" a new meaning; we are using it in an objective, not a subjective, sense. We should in reality use a new term, related to subjective "loudness" as "temperature" to "hotness."

At present there is no accepted term, though loudness-level, intensity-level, energy-level, sensation-level, and reftone (reference tone) level are used.[1]

The account just given of the decibel scale would seem to limit it to one pitch (and quality) at a time. Practical methods have, however, been found whereby it can be used for sounds of different pitch (and even different quality). Moreover the decibel scale can be applied to the power ratio in amplification of a sound, for in this case P/P_0 is equal to the ratio, I/I_0, of the intensities (as heard under the same circumstances).

571. Pitch.—The sound of a toy-whistle is of high pitch or shrill, that of an automobile horn is of medium pitch, while the tones of a church bell are deep or of low pitch. The physical cause of these differences of pitch may be shown by comparing the different sounds that can be produced by drawing a card along the teeth of a comb. If it be drawn slowly, a sound of low pitch will be heard, but if it be drawn as rapidly as possible, the sound will be of high pitch. Every time the card slips off one tooth and strikes the next an impulse is given to the air. The more numerous these impulses per second, that is, the more air-waves are started per second, the higher the pitch of the sound will be. Similar results are produced when a machine saw cuts a board. The rise and fall of the pitch of the sound is due to the variations of the speed of the saw. Savart's toothed wheel for showing the cause of the pitch of sounds illustrates the same principle.

If a phonograph be driven at lower than its normal speed, the pitch of every note will be lowered in correspondence with the decrease of frequency of the impulses imparted to the membrane.

A tuning-fork, with a small spring stylus attached to a prong, can be made to inscribe its vibrations on a revolving lamp-blacked drum or on a lamp-blacked sheet of glass, drawn beneath the fork. It is found that, for a certain speed of drum or glass, the number of waves recorded is greater the higher the pitch of the fork. When this experiment is made with sufficient care, the frequency of the fork can be accurately measured.

From the above we conclude that **the pitch of a sound depends on the frequency of the vibrations in the sound waves.** Now we have already seen that there is a simple relation between frequency and wave-length in a medium; they vary in inverse ratio, that is,

[1] Why not *belage*, like voltage?

the longer the waves the less the frequency. Hence we may also say that *the pitch of a sound in a given medium depends on the length of the sound wave.*

It must not be concluded from the above that a regular succession of air-waves produces the sensation of sound, no matter what the frequency. The rapid vibrations of the wings of a mosquito produce a sound of high pitch, and the slower vibrations of a humming-bird's wings produce a sound of low pitch; but no note of definite pitch is produced by the flight of a swallow. In fact, to produce sounds of definite pitch, the frequency of the waves must not be less than about 20 per second. On the other hand, when their frequency exceeds about 20,000 per second, air-waves do not produce the sensation of sound at all, though their existence may be shown by their action on a sensitive flame and by other

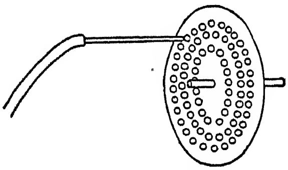


Fig. 481.—Disk siren.

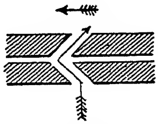


Fig. 482.

means. The upper limit of frequency is different from different persons, and is lowered by advancing age and by disease of the ear.

The siren is an instrument for producing sounds of definite pitch by a succession of puffs of air, following one another in rapid succession. In the simplest form (Fig. 481) a circular disk, with circular rings of holes, is driven at a high speed, and a puff is produced every time a hole comes opposite the end of a tube, through which air is driven under pressure. When the frequency of succession of the puffs is great enough to produce a note, the pitch can be raised by increasing the speed of the disk or by transferring the tube to a ring in which there is a larger number of holes. In the more complete form, the rotation is produced by the compressed air as it escapes from a box, the cover of which is a disk (Fig. 482) with a ring of holes corresponding to those in the rotating disk. The holes in the rotating disk and those in the fixed disk slope in opposite directions, and the jets of air impinge obliquely on the sides of the holes of the rotating disk, thus causing it to rotate. The frequency of the note is found by means of a suitable speed counter, geared to the rotating disk.

572. Doppler's Principle.—When an observer is in motion toward a source of sound, the pitch of the note heard is higher than when he is at rest. If the hearer is in motion away from the source, he hears a lower note than when he is at rest. Similar results follow when the source is in motion toward or away from the observer. The pitch of the gong of a fire engine or of the whistle of a locomotive is higher when the source is approaching the hearer than when it has passed and is receding.

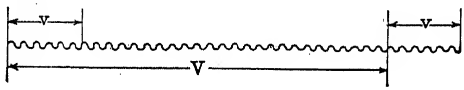


FIG. 483.

When the observer is in motion toward the source, he receives more waves in each second than when he is at rest. The additional waves received are those which occupy the distance, v, which he traverses in a second, and, if λ is the wave length, these are v/λ in number. If V be the velocity of sound and n the frequency of the source, $V = n\lambda$. Hence the increase in frequency heard is $\frac{v}{V}n$, and if the frequency of the note heard is n',

$$n' = n\left(1 + \frac{v}{V}\right)$$

For the case in which the hearer is in motion away from the source, the sign of v must be reversed.

When the source is in motion toward the hearer, the effect is a shortening of the wave length, for the source is following after the approaching waves, and the crests therefore come closer together. If the frequency of the source is n and its velocity is v, during each vibration it travels a distance v/n, and each wave length is shortened by this amount. Hence the wave length of the sound heard is not $\lambda = \frac{V}{n}$ but $\lambda' = \left(\frac{V}{n} - \frac{v}{n}\right)$ and the frequency n' of the note heard is

$$n' = n\left(\frac{V}{V - v}\right)$$

If the source is receding, the sign of v must be reversed.

The two expressions for n' do not differ appreciably if v is small compared with V.

If a vibrating tuning fork on its sounding box be moved rapidly toward a wall or blackboard, an observer will hear two notes of different pitch. One is the note heard directly from the receding tuning fork and is lowered in pitch by the motion. The other note is due to the waves reflected from the wall, and this is raised in pitch. The interference of these two notes produces beats (§578).

573. Scale of Musical Sounds.—The difference of pitch of two sounds is called the *interval* between them. In the practice of music these intervals are learned by ear, but in scientific work an interval is stated by giving the ratio of the frequency of the higher sound to that of the lower. Certain intervals have received particular names. Thus the interval between two sounds whose frequencies are as 2:1 is called an interval of an *octave.*

A *musical scale* consists of a chosen succession of notes, extending over an octave. The most familiar scale (the major diatonic) is usually learned by the syllables *do, re, mi, fa, sol, la, ti, do.* It may be started at any pitch, but in any case the succession of intervals is as given in the first line of the table below. It will be noted that this contains two short intervals of 16/15 and five long intervals of 9/8 or 10/9, which are nearly equal.

	9/8		10/9		16/15		9/8		10/9		9/8		16/15	
C		*D*		*E*		*F*		*G*		*A*		*B*		*C'*
24	:	27	:	30	:	32	:	36	:	40	:	45	:	48
264		297		330		352		396		440		495		528
256				320				384						512

Notes of definite pitch are named from letters of the alphabet, with primes or subscripts for the octave. For the "middle octave," as it is called, (*C* being "middle *C*" of a piano), the letters are as indicated in the second line of the table. In the third line is a series of numbers, such that the ratio of any two is the number that measures the interval between the corresponding notes. These we shall call the "proportional numbers" of the scale. The next line contains the actual frequencies of these notes, according to the common method of tuning orchestral instruments. In this *A* is taken as 440, but it varies somewhat in different orchestras. It may be noted that these numbers are equal to the proportional numbers multiplied by 11. The last line contains the frequencies of these notes as it has been for a long time the custom to use them

in scientific work. These are equal to the proportional numbers multiplied by $10\frac{2}{3}$.

Some other intervals in the scale are named from the order of the notes reckoned from the first. Thus the interval from the first to the fifth, that is, from *do* to *sol* or from any C to the G next above it, is called an interval of a *fifth.* From the proportional numbers we find that an interval of a fifth is equal to 36/24 or 3/2. It will readily be seen that the intervals from E to B and from F to C' (octave of C) are also intervals of a fifth (3/2). The interval C to F is called an interval of a *fourth* and is equal to 4/3, as are also $D - G$, $E - A$, $G - C'$. The intervals $C - E$, $F - A$, $G - B$ are intervals of a *third,* equal to 5/4. $C - D$, $D - E$, $G - A$, $A - B$ are all called *whole tone* intervals. The intervals $E - F$ and $B - C'$ are much smaller and are called *semi-tone* intervals. To make it possible to start the scale at any pitch, intermediate notes, named from these same letters, but "sharped" or "flatted," (the black keys on a piano) are used. This makes twelve intervals between successive notes. In a piano these are all made equal, each being $2^{\frac{1}{12}}$. In this "equal-tempered" scale all intervals (except the octave) are slightly out of tune, but the difference from a "just" scale is too slight to cause discomfort.

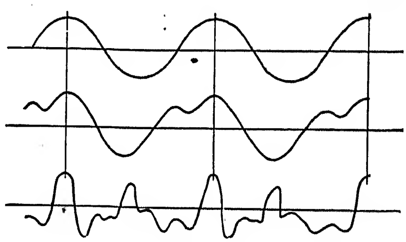


Fig. 484.—Different forms of waves.

574. Quality.—Two musical sounds of the same pitch and loudness may seem to the ear to be quite different. If the note C be sounded on a violin and on an organ, and be also sung, the ear will recognize the source at once. This difference we call a difference in *quality.*

On what does quality depend? We had no difficulty in connecting loudness with the amplitude of the wave-motion, and pitch with the frequency. Evidently, quality must depend on the only other property in which air-waves differ, namely, the wave-form, as shown by curves representing the waves. Figure 484 shows three waves, of the same amplitude and length, but of different forms. The upper is what we have called a simple harmonic wave, and is the form of wave given off by a tuning-fork or an organ-pipe, when sounded very softly. The second is more like the wave from a violin string, and the third like the wave from the

human voice, when singing the vowel *ah* (Miller). But voices differ greatly, and two voices singing the same note produce waves of somewhat different form. The same is true of waves emitted by a violin.

575. Harmonics.—Notes with frequencies 2, 3, 4, etc. times the frequency of another note are called *harmonics* of the latter. Thus, if the frequency of C is 256, the *first harmonic* of C is C', which has a frequency of 512 and is an octave above C. The *second harmonic* of C is a note with a frequency of 768 and since 768/512 is the same as 3/2, the second harmonic of C is a fifth above C', that is, it is G'. The third harmonic of C has a frequency of 1024, and since 1024/512 is the same as 2/1, it is an octave above C' or two octaves above C. Hence it is C''.

Briefly stated, notes whose frequencies are $2N$, $3N$, $4N$, etc., are called the harmonics of the note whose frequency is N, and the latter is called the *fundamental* of the harmonics.

If we now sound together such a series of notes, either with tuning-forks or on a piano, it will be found that (up to the seventh in the series) they combine together so harmoniously that it may be difficult to hear them separately. Thus the term harmonic is very appropriate.

576. Overtones.—If the sound from a large bell be listened to attentively, it will be found that it consists of several sounds, of different pitch. The deepest of these is called the fundamental of the bell, and the others are called *overtones* of the bell. The intervals between these overtones may be quite different on different bells. In some cases the combination is such that the sound of the bell is pleasing. Other bells give an unpleasing or harsh combination. A note produced on a violin by a beginner is apt to consist of an unpleasant combination of a fundamental and overtones, while an accomplished violinist has learned how to produce an agreeable combination.

The difference between the terms harmonic and overtone should be noted. Harmonics refer to notes, however they may be produced. When we are given the frequency of a fundamental, we can at once calculate the frequencies of its harmonics by multiplying by 2, 3, 4, etc. Overtones refer to a particular instrument. An overtone of an instrument may not be a harmonic of the fundamental, but we shall see that the overtones of some instruments are harmonics of their fundamental tones.

577. Elementary and Compound Sounds.—Most people can, without any special training, distinguish a single sound of definite pitch from a mixture of such sounds. Thus, a tuning fork emits a single sound when it is struck gently. When it is struck violently, or when two tuning forks are sounded at the same time, the ear can usually tell us that a mixture of sounds is heard.

A well-trained ear is capable of doing more than this. For example, it can detect that, when a single piano string is struck, the sound produced is not single, but consists of a fundamental and various harmonics. Thus the ear can do for musical sounds what the chemist can do for chemical compounds, namely, resolve them into their elements. The simple or elementary sounds of music are those in which the ear can detect no mixture of sounds. These are called *pure tones.* This at once suggests another question. What is the difference of the wave form of a pure tone and that of a mixed note, consisting of a fundamental and harmonics?

Now we have already seen that there are means by which a vibrating body can be made to record the form of its vibrations (§571). By these it is found that **the vibrations of a body emitting a pure tone are simple harmonic motions, and the sound waves it emits are simple harmonic waves.** We have also seen that simple harmonic waves of different length, when added together, produce complex waves, which may differ markedly from the simple harmonic form. Thus, harmonics present with a fundamental alter the form of the wave, and it is this wave form that determines the quality of the sound heard.

578. Beats Between Sounds.—When a white key and an adjacent black key near the bottom of the keyboard of a piano are struck at the same time, a distinct throbbing of the sound can be heard. The throbs are slow and can almost be counted. When the same is tried higher up on the keyboard, the throbs are more rapid. Throbs produced by two sounds of nearly the same pitch are called *beats.*

If two tuning forks of nearly the same pitch and mounted on their sound boxes be thrown slightly out of unison by attaching a small piece of wax to a prong of one, and if they be then sounded strongly by a bow, beats slow enough to be counted will be heard. When a larger piece of wax is used the difference of pitch of the forks is increased, and the frequency of the beats also increases.

In fact, in all cases, **the frequency of the beats between two notes is equal to the difference of the frequencies of the notes.**

Since beats are due to two wave trains, of different wave lengths, coming to the ear at the same time, they are the result of what we have called *interference* of waves (§247). Figure 152 may be taken as representing the production of beats between two trains of sound waves.

When beats between two tones are sufficiently rapid, they coalesce and form a distinct beat tone, which may, however, be very weak. The frequency of the beat tone is the difference of the frequencies of the separate tones, and when it is distinctly audible it is usually much lower than either component. When careful attention is paid, the beat tones between two piano strings or two violin strings can be clearly heard, and, faint though the sounds may be, they have a distinct effect on the musical quality for trained ears. Beat tones may, if of too low a pitch to awake resonance in the instrument, which is the case when the beating tones are on the lower strings of a violin, be an effect purely in the ear, while the instrument itself may resonate to higher beat tones. Helmholtz showed that all discords between two notes are due to beat tones between fundamentals or harmonics.

579. Nature of Vocal Sounds.—A vowel is a sustained sound of a variable pitch, produced by holding the vocal cords and the resonating cavities of the mouth and throat in definite configurations. Consonants are explosive noises, produced by the changes in the vocal organs preceding or following a vowel sound. Vowels, uttered by the same voice at the same pitch, differ in quality. Figure 484 shows a record obtained by Miller for the vowel *ah* in father at a pitch of 182 by means of his phonodeik (§562). The precise cause of the difference of quality of different vowels has been a matter of long dispute. The first and natural supposition was that a certain vowel consisted of a definite combination of a fundamental and its harmonics in fixed proportions, whatever the pitch of the fundamental might be. If so the curve of Figure 485 would be of the same form (for the same voice) whatever the pitch of the fundamental might be. This is contrary to results obtained by Miller and others. Helmholtz maintained that, in a particular vowel sound, there was a definite overtone characteristic of the vowel and having the same frequency, no matter what the pitch of the fundamental might be. Miller found that, at whatever pitch (between 129 and 259) the vowel *ah* was sounded, 60 per cent. of the energy was concentrated in an overtone of about 920, which was nearly but not quite constant. When a phonograph is driven at less than its normal speed, the quality of a vowel in a singer's record is altered. This is contrary to the first view, since the change of speed does not change the *relative* pitch and energy of the combination of fundamental and overtones. The overtone characteristic of a vowel is the tone to which the mouth cavity resonates. It can be heard by whispering the vowel.

580. How We Hear.—Some parts of the process of hearing are not yet well understood; so the following brief account will be

confined to features about which there seems to be general agreement.

The organ of hearing is, for convenience of description, regarded as including three parts, the outer ear, the middle ear and the inner ear. The latter two are, for clearness, exaggerated in size in Fig. 485. The outer ear consists of a roughly cup-shaped sound collector (not of much use in man) and a passage (meatus), about an inch long, that leads to the middle ear. This passage ends in a membrane *A*, usually called the *eardrum*. The middle ear, which begins at A, contains air and a chain of small bones, called, from their peculiar shapes, the "hammer" (malleus) *B*, the "anvil" (incus) *C*

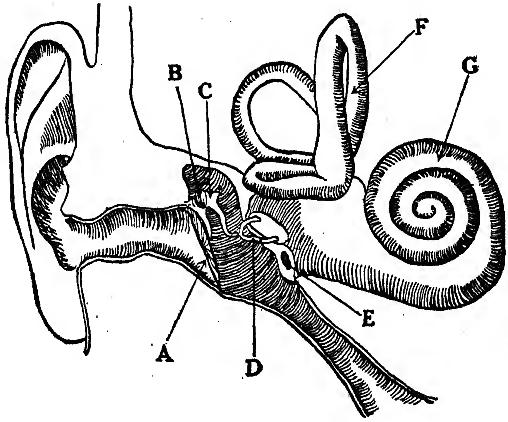


FIG. 485.—Structure of the ear. (From Weld and Palmer.)

and the "stirrup" (stapes) *D*. They link the eardrum to the *labyrinth*, as the inner ear is otherwise called. Two small parts of the wall of the labyrinth are flexible membranes, the *vestibule* or oval window, to which the stirrup is attached, and the round window E.

The labyrinth is a very irregular chamber, filled with a liquid (lymph) and consisting of two parts that serve quite different purposes. The upper part, called the *semicircular canals F*, acts as a kind of spirit level to aid in keeping the body upright. The lower part, the *cochlea* (snail-shell) *G*, contains a membrane on which the end of the auditory nerve is spread out. This membrane is a partition that divides the cochlea, for nearly its full length (about an inch), into two parts, leaving a small opening at the tip of the cochlea through which slight surgings of the liquid between the

two parts can take place. But just how this most remarkable of all acoustical mechanisms sorts out and transmits to the brain center the immense variety of details of the complex sound patterns that fall on the ear is almost wholly unknown.

SOURCES OF MUSICAL SOUNDS

581. Musical Instruments.—The number of musical instruments is so great that only a few can be referred to here. They may be most simply classified according to the kind of body that is started vibrating in order to produce the sound.

Vibrating cords are used in violins, pianos, harps, etc. In an orchestra these are called *stringed instruments.*

Vibrating columns of air are used in organs, flutes, clarinets, etc. These are called *wind instruments.*

Vibrating rods, plates, bells, and membranes are used in what are called *percussion instruments.* They are sounded by striking.

582. Vibrations of Cords.—We have already considered cases of vibrations on cords when the vibrations are slow enough to be followed by the eye (Figs. 137, 160). When such waves are continually reflected from both ends, the cord can divide up into vibrating segments or stationary waves. Each such segment is half of a wave-length in length. We shall now consider vibrations on a cord when they are rapid enough to produce sound waves, but there is no essential difference between the two cases.

When a cord stretched between two supports vibrates, it moves to and fro between two opposite extreme positions. The forms of the cord in the two extreme positions can be found by photography and in other ways, and they are found to depend on the way in which the cord is started. The simplest case is when the cord is very gently bowed at the middle. In this case the form is that of half of a simple harmonic wave (Fig. 486, *A*), and, as might be expected, the note produced is a pure or elementary tone. The wave-length is therefore $2l$, where l is the length of the cord. Now the velocity of waves equals the product of frequency and wave-length or $v = n\lambda$. Hence

$$n = \frac{v}{2l}$$

and, since the value of v is $\sqrt{T/m}$ (§250), T being the tension and m the mass per unit length of the cord,

$$n = \frac{1}{2l}\sqrt{\frac{T}{m}}$$

From this we see that the frequency is inversely proportional to the length. By shortening the "string" a violinist produces a higher note.

The frequency is also proportional to the square root of the tension. To tune a violin string to the proper pitch, the tension is increased or decreased by turning a peg on which one end of the string is wound.

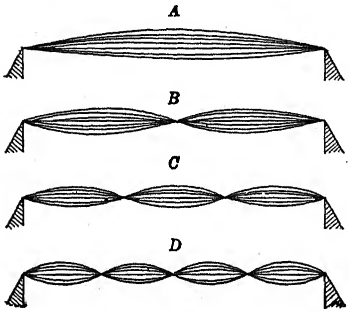


Fig. 486.

Finally, the frequency is proportional inversely to the square root of the mass of unit length. A violin has four "catgut" (sheep's intestine) strings, the heaviest being used for the lowest notes and the lightest for the highest notes, though a fine steel wire is often used for the highest-string.

The vibration of a violin string is maintained by the bow, which alternately grips and slips on the string. The sound is not emitted directly by the strings, but by the resonating body of the instrument.

583. Overtones of Cords.—A cord can also divide into 2, 3, 4, etc., vibrating segments, as we have already seen in Fig. 137, and this is true whether the vibrations are too rapid to be followed by the eye or not. To produce vibrations with two segments on a cord, touch it lightly at the center and bow one-half (Fig. 486, *B*). It will be seen that both halves vibrate, and the ear will hear a note that is higher by an octave than the fundamental. This is the first harmonic of the fundamental. If the finger be one-third of the length of the string from one end, and either the shorter or the longer part of the string be bowed, the note will be the second harmonic of the fundamental, and so on. In this way the violinist can produce a great variety of harmonics on the various strings.

Thus the overtones of a cord are harmonics of the fundamental. To explain why this should be so, let us suppose that in any case the number of segments is N. Then, if l is the length of the cord, the length of each segment is l/N, and the wave-length is therefore $2l/N$. From this and the general relation $v = n\lambda$, we get

$$n = \frac{v}{\lambda} = \frac{vN}{2l}$$

Now l is the constant length of the cord, and v is also constant so long as the tension of the cord is unchanged. Hence, if we give to N values 1, 2, 3, etc., we see that the frequencies are also as 1, 2, 3, etc., the tones being the harmonics of the fundamental.

584. Complex Vibrations of Cords.—While we have described separately the different ways in which a cord can vibrate, they can, in reality, take place at the same time. When a cord is bowed at one-fourth of its length from one end, a good musical ear can detect both the fundamental and the first harmonic. Or if, while it is vibrating in this condition, it be lightly touched at the middle, the fundamental will cease, but the first harmonic will be heard to continue for a moment. In a similar way, we can show that other harmonics are present.

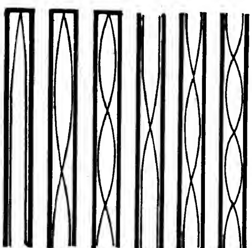
The particular overtones that are present in the complex vibrations of a cord depend chiefly on where it has been struck or bowed, for it is evident that *no form of vibration that would require the point struck to be a node can be present.* Some of the possible combinations are more pleasing than others. The most pleasing combination is obtained when the cord is struck at about one-eighth from one end, and this is the actual practice in the construction of the piano, and also, to some extent, in the bowing of the violin. It nearly eliminates $7N$ and $9N$ which are not in harmony with the lower harmonics of N, and $8N$ is sacrificed.

The explanation of the coëxistence of different forms of vibration depends on the fact that waves of different length can travel along the cord at the same time. Each such train, being reflected at the ends, produces its own system of nodes and vibrating segments.

The form of the resultant wave on the string evidently depends on the particular combination of vibrations present, and for each combination there is a distinct characteristic quality of the complex sound heard. This is another proof that the quality of a sound depends on the form of the wave that causes it.

585. Vibrations of Air Columns.—Air in a tube, open at one end or at both ends, can be made to vibrate and emit a musical sound, by blowing across an open end. Blowing with different degrees of strength will produce notes of different pitch. Gentle blowing will produce the lowest note, and we shall consider this first.

Let us first suppose that the tube is closed at one end, and, for brevity, let us call this a stopped tube. The vibrations of the air column are stationary waves (Fig. 487, *a*). Compressions, started by the blowing at the open end, travel to the closed end, and are there reflected, so that the motion at any point in the tube is due to the superposition of two trains of waves, travelling in opposite directions. The closed end is a place of no motion and is therefore a node of the stationary waves. The open end is a place of the greatest freedom of motion and is therefore the middle of a loop or vibrating segment. Now, the distance from a node to the middle of a vibrating segment is one-fourth of a wave-length, λ. Hence, if l be the length of the tube,



a b c d e f

FIG. 487.

$$\tfrac{1}{4}\lambda = l \text{ or } \lambda = 4l$$

If the tube is open at both ends, it is found that there is a node at the middle (Fig. 487, *d*), while the ends are middles of loops. Hence it is readily seen that

$$\tfrac{1}{4}\lambda = \tfrac{1}{2}l \text{ or } \lambda = 2l$$

From this we see that a stopped tube and an open tube of the same length produce waves with lengths as 2 to 1, and the notes have, therefore, frequencies as 1 to 2, since frequency varies inversely as wave-length. The note of the stopped tube is therefore an octave below that of the open tube.

Strictly speaking, an open end is not exactly the middle of a loop, for a condensation, returning to the open end, does not reach full freedom of expansion until it has passed out a short distance. It has been found that, for a tube of circular section, the middle of the loop is, in reality, about .6 of the radius of the tube beyond the open end.

586. Resonance of Air Columns.—If, over the open end of a stopped tube, we bring a vibrating tuning fork of the same frequency as the lowest note which the tube will emit, *resonance*, or response of the tube to the fork, will take place, and the note of the fork will come out strongly. A tube of variable length can be tuned exactly to the pitch of the fork, and for this purpose a tube closed at the lower end by a column of water that can be raised or lowered is suitable. An open tube with an extension piece that slips over the end of the tube can be tuned to a fork. If a

stopped tube and an open tube be tuned to the same fork, the former will be half as long as the latter.

The interaction between the fork and the tube is similar to that between the hand and a swing, when the latter is being started. The forward and backward motions of the hand must be timed to agree with the corresponding natural motions of the swing. Similarly, the fork must complete one vibration in the time of one vibration of the air column, that is, in the time in which the sound travels one wave length, which, as we have seen, is four times the length of the stopped tube and twice that of the open tube.

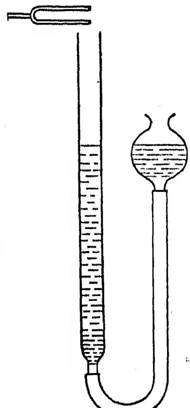


FIG. 488.—Resonating tube.

For many purposes it is convenient to mount heavy tuning forks on resonating boxes, each of the same fundamental pitch as the fork (Fig. 490). While the fork and the box agree in their fundamental tones, they differ as regards their overtones, and the note emitted is much louder and purer than the fork alone can emit, but it does not continue as long after the fork has been struck.

Cavities of other forms can also resonate to definite notes and emit the same notes when excited. A spherical resonator (Fig. 489), with an opening at B for admitting the sound and another at C to be applied to the ear, has been found especially useful. Such a resonator is highly selective in its resonance, because its overtones

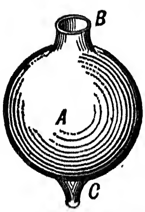


FIG. 489.—Spherical resonator.

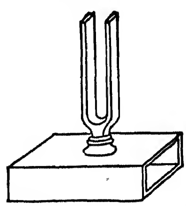


FIG. 490.—Tuning-fork on resonating box.

are high and not harmonics. Helmholtz, an eminent German physicist (1821–1894) used a series of spherical resonators for analyzing complex sounds. From this an instrument for the measurement of sound intensity has been developed by replacing the ear at C by a hot wire (or grid). Its resistance is affected by the small alternating air stream through C and can, of course, be measured

with high accuracy. Similarly a Rayleigh disk (§562) can be used at C; and a vibrating reed, carrying a small mirror, as in Miller's phonodeik (§562), can be mounted at B.

587. Organ Pipes.—An organ pipe is a tube (Fig. 491), the air in which is maintained in vibration by a jet of air from a wind chest. The jet is forced through a narrow slit at the mouth of the pipe, and strikes against a sharp edge, which borders an opening on one side of the pipe. The farther end of the pipe may be open or stopped, but the end at the mouth must be regarded as open. The first effect of the jet is to start either a condensation or a rarefaction in the pipe, depending on whether the jet is on the whole directed more to the inner or the outer side of the sharp edge. Thereafter the course of events is determined by the return of the pulse after reflection from the farther end. The arrival of a condensation forces the jet outward, and it then, by its suction, reinforces the rarefaction which succeeds

FIG. 491.

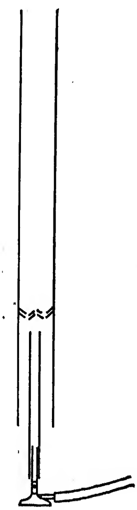


FIG. 492.

the condensation, and the rarefaction travels up the pipe. When a rarefaction arrives, the jet is forced inward and reinforces the condensation that follows. Thus the pitch is controlled by the pipe, and the energy required is derived from the jet.

A stopped pipe is tuned by adjusting a movable plug that closes the stopped end. An open pipe usually has a small hole near the open end, and the pipe is tuned by adjusting a small strip of metal that partly closes the hole.

An interesting illustration of another method of supplying the energy of vibration of a sounding tube is shown in Fig. 492. A long thick metal tube has a sheet of wire gauze inserted to about one-fifth of its length from one end. If the gauze be heated by a Bunsen burner (with a long extension-tube), the tube will sound

very loudly when the burner is removed. The whole explanation is complex, but the supply of energy is readily explained. A condensation, coming to the heated gauze, removes heat more rapidly than a rarefaction does. Hence the energy of each condensation is reinforced by the supply of energy received from the heated gauze.

Knipp's "singing tube" (Fig. 483) is a very ingenious development of the preceding (of much smaller size). It consists of a tube within a tube. When the closed end of the outer tube is heated a sustained tone is emitted. Two such tubes are convenient for producing sustained beats (§578).

FIG. 493.

588. Tones of a Stopped Organ Pipe.—The closed end of a stopped pipe is always a node, and, for the fundamental tone, it is the only node, the open end being the middle of a loop. The first overtone has a second node, the open end being still the middle of a loop (Fig. 487, *b*). Hence the second node is one-third of the length of the pipe from the open end. Thus the wave-length, which is always four times the distance from the middle of a loop to the nearest node, is 4/3 of the length of the pipe. For the second overtone there are three nodes and so on. Thus it is readily seen that, if l be the length of the pipe and N the frequency of the fundamental, the series of wave-lengths and notes are

$$\begin{array}{lll} 4l, & 4l/3, & 4l/5, \ldots \\ N, & 3N, & 5N, \ldots \end{array}$$

It will be noticed that the overtones are harmonics of the fundamental, but those of frequency $2N$, $4N$, etc., are not produced by a stopped pipe.

Except when blown very softly, a pipe produces one or more overtones along with the fundamental, and the number and intensity of these determine the quality of the complex sound. The absence of the harmonic $2N$ gives the stopped pipe its somewhat dull tone.

589. Tones of an Open Organ Pipe.—Both ends of an open pipe are mid-points of loops. At least one node is required for stationary waves, and, when an open pipe sounds its fundamental tone, this node must evidently be at the middle of the pipe. For

the first overtone there are two nodes, one at one-fourth of the length of the pipe from one end, and the other at an equal distance from the other end (Fig. 487, *e*). For the second overtone there are three nodes, and so on. Thus it is readily seen that the wave-lengths and frequencies are

$$\begin{array}{lll} 2l, & 2l/2, & 2l/3, \ldots \\ N, & 2N, & 3N, \ldots \end{array}$$

The bright quality of the open pipe is due to the overtone $2N$.

590. Longitudinal Vibrations of Rods.—A rod, clamped at its middle point, can be made to vibrate longitudinally by stroking it with a rosined glove. Since both ends are free and the middle is fixed, the fundamental vibrations are similar to those of an open organ pipe, and the wave-length is twice the length of the rod. The overtones follow the same law as those of an open organ pipe. The first overtone may be produced by clamping the rod at points one-fourth of the length from the ends (as in Fig. 497), and so on.

591. Transverse Vibrations of Rods.—A rod or strip of metal, clamped at one end, can readily be set into transverse vibrations by a blow at the free end (Fig. 494). Of two rods of the same material and cross-section, the shorter vibrates more rapidly than the longer, and, if the vibrations are rapid enough, they produce musical notes. The fixed end is, of course, a node, but overtones with two or more nodes can also be produced. The frequencies of these are not in simple ratios to the frequency of the fundamental. Hence the overtones are not harmonics. Such vibrations of strips of wood or metal are used as *reeds* in many musical instruments, such as the clarinet, oboe, bassoon, and reed organ. The reed is placed at the end of a pipe that resonates to the vibrations of the reed.

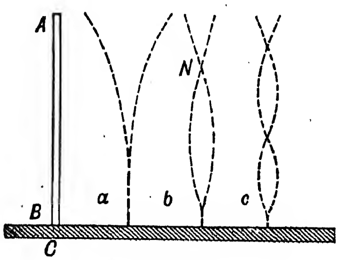


FIG. 494.

A rod, clamped at the middle, will vibrate transversely, with a node at that point. A tuning fork consists essentially of such a rod, with the ends turned parallel. The overtones of a tuning fork are not harmonics of the fundamental, but, if the fork is sufficiently thick and is not struck too violently, the overtones, besides being very high, are so weak as not to produce disagreeable effects.

Large metal tubes, vibrating transversely, are sometimes used in operatic music to imitate deep-toned bells.

592. Vibrations of Plates.—A square or round plate of metal, the middle of which is screwed to the end of a rod, can be made to vibrate in a great variety of forms by stroking with a bow. The nodes, or lines of no motion, are beau-

tifully shown by fine sand on the plate. In the fundamental mode of vibration there are only two of these lines, but in the higher modes of vibration a large number, straight or curved, appear (Fig. 495). A square plate may be regarded as a set of rods parallel to either of the pairs of parallel sides. The combination of any partial vibration of one set with any one of the other set accounts for the great number of possible forms of vibration of the plate.

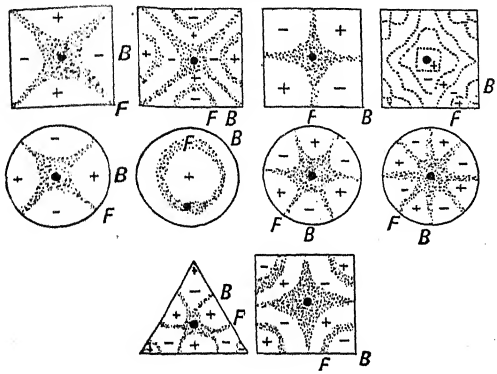


FIG. 495.

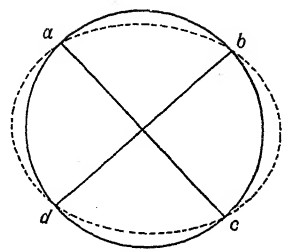


FIG. 496.

593. Bells.—A bell may be thought of as a round plate bent to a cup shape (Fig. 496). The fundamental mode of vibration is like that of the plate, with two nodal lines; but, when the bell is struck, many overtones are produced along with the fundamental, and some of these are much stronger than the fundamental. The "strike note" or note that the ear judges to be the pitch of the bell, is usually an octave below the fourth overtone and is not itself an overtone. Possibly it is the difference (or beat) tone between the fourth and sixth overtone.

VELOCITY OF SOUND.

594. Resonating Tube Method.—From the length of a tube that resonates to a tuning fork of known frequency, the velocity of sound in the gas in the tube can be found by means of the relation $V = n\lambda$. For this purpose a glass tube open at one end and containing an adjustable column of water is convenient. From the account of the vibrations of a stopped organ pipe (§588), it is seen that, if the tube be of sufficient length, it can resonate when the length of the column of gas is $\lambda/4$, or $3\lambda/4$ or $5\lambda/4$, etc., where λ is the wave length of the sound of the tuning fork. When it resonates with more than one node, the distance between two consecutive nodes is accurately $\lambda/2$, while the distance from the upper node to the open end is only approximately $\lambda/4$. Hence the best value of the wave-length and velocity is obtained from the distance between nodes.

595. Kundt's Dust-tube Method.—This method differs from the preceding in that a metal rod, vibrating longitudinally, is used instead of a tuning fork, and, since the pitch is very high, the air column divides into numerous vibrating segments. These segments are clearly shown by cork dust, which gathers at the nodes. One end of the rod projects into the tube and carries a light disk, of slightly smaller diameter than that of the tube. It

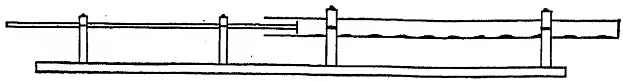


FIG. 497.

is convenient, for stability, to clamp the rod at two points, each at one-fourth of the length of the rod from one end. The wave-length of the waves in the rod is then equal to the length, L, of the rod. The wave-length of the sound in the gas is twice the distance between two adjacent dust heaps, or $2l$. Let V be the velocity of the waves in the rod, and v the velocity of sound in the gas. Then, if n be the frequency of the vibration, which is the same for both, $V = nL$ and $v = 2nl$. Hence

$$\frac{v}{V} = \frac{2l}{L}$$

If we assume the velocity of sound in air to be known, we can, from this relation, find the velocity of the waves in the rod and also the velocity of sound in any other gas introduced into the tube. The method also lends itself to the study of the effect of change of temperature on the velocity of sound, and it has been modified so as to make it suitable for finding the velocity of sound in a liquid.

PRACTICAL APPLICATIONS

596. Acoustics of Halls.—In some halls the "hearing" is satisfactory, in others unsatisfactory. A full treatment of this extensive subject is impossible here, but a few points may be considered.

Sound reaches an auditor in a hall, not only by waves coming directly from the source, but also by waves reflected by walls (including in this term all reflecting surfaces), and the paths usually differ in length. The result may be an increase in loudness, but more or less confusion of sound also follows. A speaker utters two or three syllables per second, and a musical instrument may

emit several notes in a second, and reflection causes an overlapping of the separate sounds. The two chief defects resulting may be classified as *echo* and *reverberation*. A separate echo is due to

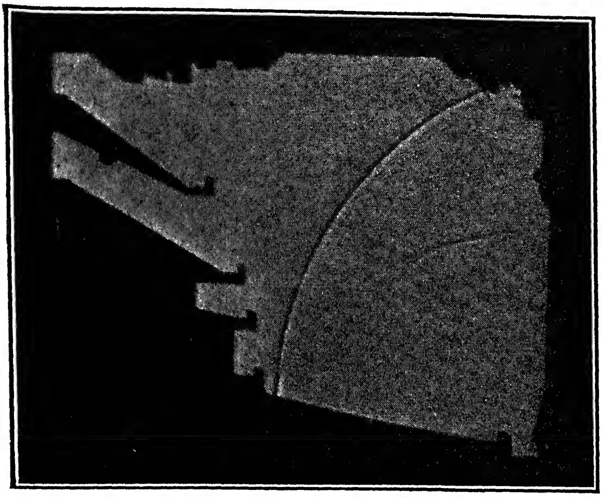


FIG. 498*A*.—(Sabine, *American Architect*.)

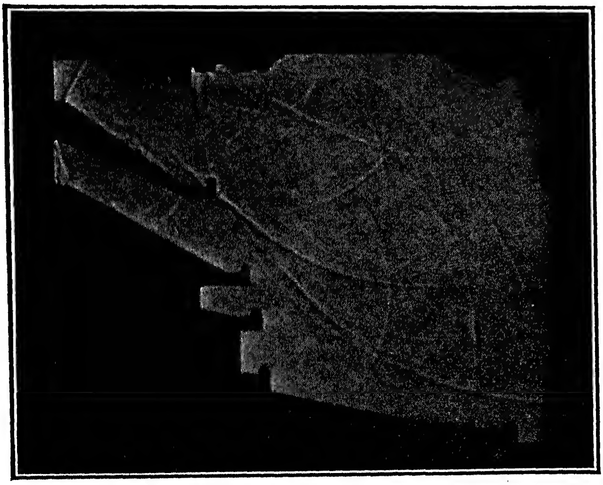


FIG. 498*B*.—(Sabine, *American Architect*.)

some particular reflecting surface and may be reduced by alterations in the surface. Reverberation consists in a prolongation of the sound in all parts of the hall, due to repeated reflections from all walls, somewhat as a room (with white walls) may be illuminated

throughout by light entering by a single small window. Reverberation is mostly a drawback for speech, but it cannot be eliminated without reducing loudness to an undesirable extent. For music a certain amount of reverberation is an advantage.

Figure 498 shows the history of a single wave of sound, started from the stage of a theater and reflected by various surfaces. In *A* we have the direct wave, .07 of a second after it has started, together with several reflected waves that will cause weak echoes. *B* shows a part of the direct wave, .07 of a second later, and a complex system of reflected waves. One of these is coming to the fronts of the balconies as an echo from the ceiling. This echo was eliminated by changing the form of the ceiling. The figures are from photographs by W. C. Sabine, American physicist (1868–1919) who was the first to place architectural acoustics on a scientific basis (1900). The method used in getting the photographs was that referred to in §564, a small model of the theater being used for the purpose.

597. Remedy for Reverberation.—If the walls of a hall did not reflect sound, there would be no reverberation. When it is present to an undesirable extent, its amount may be decreased by covering some part of the walls with soft materials that reflect little sound. It is usually not a matter of importance on what wall or part of a wall the deadening material is placed. This is due to the fact that the sound starts out from the source in all directions and travels about 1140 ft. in a second. This distance is so great, compared with the dimensions of an ordinary hall, that the rays are, in general, reflected many times in a second, and as much approximately falls on each square unit of area, wherever it may be situated. Yet there are places in some halls where more or less than the average amount of sound reaches the walls, and, for such special cases, a special study of the hall is necessary.

598. Duration of Residual Sound.—Since reverberation is due to the continuance of a sound after it has been produced, the extent of the reverberation in a hall can be estimated by measuring the length of time a sound is heard in the hall after the source has ceased emitting waves. Sabine studied this question by means of an organ pipe, having a frequency of 512 and blown under a definite pressure, together with a chronograph for recording the length of time the residual sound was heard. He found that the residual sound was heard for a measurable number of seconds

and died away according to an exponential law. From the data obtained it was possible to calculate the time required for any specified decrease in intensity, and it was convenient to define "the duration of residual sound" in a hall as the time required for the intensity to fall to one millionth (10^{-6}) of its original value. This we shall denote by T. While the value of T is different for notes of different pitch and such variations were studied by Sabine, it will be sufficient, in this brief account, to restrict it to the particular frequency (512) referred to. An auditorium must usually serve for both music and speech, and the requirements, as regards reverberation time, are not so different as to make this a serious disadvantage. The best reverberation time increases with the volume of the auditorium. Thus for a volume of 20,000 ft.^{3} T should be about 1.2 sec., for 200,000 ft.^{3} about 1.6 sec., and for 1,000,000 ft.^{3} about 2.0 sec.

599. Coefficients of Absorption.—An open window may be regarded as a non-reflector, that is, a perfect absorber, since all the sound that falls on it passes out. Sabine found, by experiments on the duration of residual sound in an auditorium, what area of any particular drapery would cause the same absorption of sounds as 1 sq. m. of open window, and therefrom he deduced the absorbing power of the material. If, for example, it took 3 sq. m. of a material to produce the same decrease of the residual sound as 1 sq. m. of open window, each square meter of the material must have absorbed one-third as much as the square meter of open window. But the latter absorbed all of the sound that fell on it. Hence the drapery absorbed one-third of the incident sound, and therefore its coefficient of absorption was $\frac{1}{3}$. By such experiments Sabine found the absorbing power of an average auditor and of the various articles that are commonly found in a hall.

COEFFICIENTS OF ABSORPTION

Open window	1.000
Wood sheathing (hard pine)	.061
Plaster on lath	.034
Plaster on tile	.025
Brick set in Portland cement	.025
Carpet rugs	.20
Shelia curtains	.23
Hair felt (2.5 cm. thick, 8 cm. from wall)	.78
Audience (per person)	.44
Wooden seats (per seat)	.003
Upholstered seats (per seat)	.300

These figures were for a frequency of 512. Subsequent work, with a wide variety of materials, has shown that the coefficients depend also on the frequency and on the condition of the materials.

600. Total Absorbing Power of a Hall.—Each surface absorbs in proportion to its area and coefficient of absorption. If, therefore, we multiply each area of surface of a particular kind by its coefficient of absorption and add the products we shall get the total absorbing power, A, of the auditorium, or

$$A = a_1s_1 + a_2s_2 + \cdot \cdot \cdot = \Sigma as$$

To take account of the audience in this sum, we multiply the number supposed to be present by the absorbing power per person.

Sabine found that, in general, the *shape* of a hall, of given total volume and given total absorbing power, had, to a first approximation at least, no effect on the length of duration of residual sound in the hall. The sound diffuses with great rapidity to all parts of a hall, and so becomes nearly uniformly distributed throughout. Hence the absorption must depend only on the average frequency of reflection, and this depends only on the volume.

601. Sabine's Formula.—By experiments on many halls Sabine found that the length of duration of residual sound could be expressed by

$$T = .164\frac{V}{A}$$

where V is the total volume of the hall in cubic meters, and A is its total absorbing power, areas being calculated in square meters. Strictly speaking, the velocity of sound should appear in the denominator, but it is assumed to have the average value of 342 meters per second and is included in the factor .164. When feet are used, the factor 0.164 is to be replaced by 0.050.

Sabine's formula can be applied to calculating the duration of residual sound in an auditorium, existing or contemplated, and thus its suitability for music, and approximately also for speech, can be tested. It can be applied also to remedy the defects in an existing hall by suitably varying the value of a. From the formula we can also derive immediate answers to such questions as: the effect of the presence of the audience on the "hearing" in a hall; the result of copying an existing satisfactory hall but on a larger scale; and so on. These may be left as exercises to the reader (see problem 10).

Later investigators have found that Sabine's formula, while applicable to halls of moderate dimensions and moderate damping ("live" halls), is not adequate where the damping is excessive ("dead" halls) or the volume exceptionally great (above 1,000,000 cu. ft.). For other, more complex formulae that have been found more general, special works on architectural acoustics must be consulted.

602. Other Applications of Sound.—The principles on which the directions of sound are detected by the two ears (§567) are applied in instruments for detecting the direction of sound. Instruments on the *difference of intensity* principle, others on the *difference of phase* principle, and still others using both principles, are employed.

As an aid to navigation in foggy weather, two sound receivers, on opposite sides of the hull of a vessel, serve as two ears for the vessel for hearing the sound of a submerged bell, these sounds being conveyed through tubes (or electrically) to the navigating officer on deck. The two sounds do not appear equally loud unless the vessel is pointed in the line of the source. The actual direction is ascertained by changing the direction of the vessel, just as the direction of a sound is found by turning the head.

Intruments with two receivers, mounted on an arm that can be rotated beneath the vessel, the sounds being conveyed by tubes of equal constant length, make changes in the direction of the vessel unnecessary. It is, however, not necessary to have a rotating arm, for two fixed receivers a few inches apart, will, in general, receive sounds in different phase, and these, if conveyed to the two ears by tubes of the same length, will produce in the head the same effect as sound in a corresponding direction, heard directly. By making one of the tubes of an adjustable length, the sound can be made to appear to come from directly ahead; an index, geared to the apparatus for varying the length, indicates the real direction of the source. In this way, the direction of a motor-driven surface or submerged vessel can be detected, to an accuracy of two or three degrees at a distance of several miles. The sensitiveness of the apparatus is greatly increased by using numerous pairs of receivers, properly combined.

Instruments on the last principle, and placed near the bow of a vessel, can be turned so as to hear the sound of the propeller reflected from the bottom, if the depth of the water does not greatly exceed the length of the vessel. For greater depths the time required for an echo from the bottom is observed. By these devices soundings for depth can be continuously taken while the vessel is in motion.

Supersonic "sounds" of frequencies up to 1,000,000 or more are produced by the oscillations of crystals, especially quartz. The crystals are piezoelectric (§398) and are stimulated to resonant emission by oscillating electric fields. Detectors for such sounds depend on the same principle reversed. They have remarkable properties, such as accelerating chemical reactions, starting crystallization and destroying red blood corpuscles. While originally scientific curiosities, they have been found remarkably efficient, when produced under water, as a means of detecting the nearness of icebergs in foggy weather and taking soundings while the vessel is in motion.

PROBLEMS

1. What is the ratio of the velocity of sound on a hot summer day (35°C.) to that on a cold winter day (−20°C.)?
2. A band is playing on a steamer that is travelling with a speed of 20 miles per hour. With what speed must an observer move in the opposite direction, in order that every note may be lowered by a semitone?
3. Find the pitch of the fundamental and of the first two overtones of a stopped pipe 1 m. long at 16°C.
4. What is the length of an open pipe, if the pitch of its first overtone is 512?
5. A tuning fork makes 2 beats per second with a standard fork, the frequency of which is 512. When a small piece of wax is placed on a prong of the former, the number of beats is decreased. What is its pitch?
6. A tuning fork of frequency 384 makes 2 beats per second with a vibrating string. In what proportion must the tension of the string be changed, in order that the two may be in unison?
7. A stopped pipe resonates to a tuning fork, the pitch of which is 258, when adjusted to a length of 32 cm., and also when the length is 98 cm., the temperature being 18°C. What is the ratio of the specific heats of air?
8. A glass rod is used in Kundt's experiment. If the dust piles are 5 cm. apart and the distance between the clamped points on the rod is 70 cm., what is Young's modulus for this glass, the density of glass being 2.7 gm./cu. cm.?
9. What is the depth of water under a vessel 400 ft. long, if the sound of the propeller, heard by a sound detector at the bow, seems to come from a direction making 40° with the horizontal.
10. (Adapted from Sabine). A certain hall that has been found well-suited for orchestral music is constituted as follows:

Volume of hall	11,200 cu. m.
Plaster on lath	2,206 sq. m.
Wood sheathing (pine)	235 sq. m.
Drapery	80 sq. m.
Audience and orchestra	1,600 persons

(*a*) Calculate T for this hall.

(*b*) It is proposed to copy the hall closely, but with increased dimensions, so as to accommodate 2600 people. Calculate T for this proposed hall, and criticise the proposal.

(*c*) In what proportions should the dimensions of the original hall be increased for the new hall, the additional audience being accommodated in galleries?

(*d*) What would be the effect in the new hall, as at first proposed, if half of the seats were vacant, (1) if the seats were of plain wood, (2) if the seats were upholstered?

LIGHT

By E. Percival Lewis, Ph. D.

Late Professor of Physics in the University of California

Revised[1] by F. A. Jenkins, Ph. D.

University of California

GENERAL PROPERTIES

603. Radiation.—As in the cases of Heat and Sound, the word Light has acquired two distinct meanings. The primary and more familiar one is that which is associated with the sensation of vision. Nearly all that relates to this aspect of the subject lies within the province of the psychologist. The physicist, however, generally uses the term in an objective sense, with reference to the external agencies which may excite the sensation of light if allowed to act on the eye. The visible radiation which affects a normal eye will also affect a photographic plate, a thermometer or other sensitive detector of heat. It will be found, after analyzing the radiation from the sun, electric light, or other sources with a prism, that beyond the violet and the red lie non-luminous radiations which will affect a photographic plate or a thermometer, and it has been shown that the oscillations of an electric spark between metallic terminals are accompanied by the radiation of electric waves through space. There is, as we shall see, no fundamental qualitative difference between these various radiations, and it is due merely to a special property of the eye that some of them excite the sensation of light while others do not. This is analogous to the selective resonance of a piano wire, which will respond to certain notes and not to others. Just as some ears can detect sounds of such high pitch as to be inaudible to others, some eyes can detect radiations lying somewhat beyond the limits of perception of the ordinary eye. In the following pages the whole range of these radiations so far as they are known will be considered. As a matter of convenience, the term **Light,** which strictly speaking would

[1] See footnote p. 165.

apply only to the radiations exciting the sensation of light, will often be used in a figurative sense to include the entire range of radiations which are alike in their general properties, and which were once very artificially classified as luminous, actinic, and heat radiations.

604. Sources of Light.—The best known sources of light are: the sun, the physical nature and condition of which are as yet not fully understood; solid bodies at a high temperature, such as the carbon electric arc, incandescent lights and luminous flames; and luminous gases, such as the mercury arc and neon tubes. If a piece of cold porcelain is held over the flame of a candle, lamp, or gas jet, it will become covered with finely divided carbon, while no such deposit is observed in the case of a non-luminous Bunsen or alcohol flame. This suggests that the luminosity of these flames is due to the presence of incandescent carbon particles. This idea is strengthened by the fact that when the base of a Bunsen burner is closed, the flame becomes luminous and smoky; when open, enough oxygen is admitted to combine with all the carbon set free by the dissociation of the coal gas, and the flame is then non-luminous. The substances formed are gases, carbon oxides, and these gases cannot be made luminous by the highest temperatures attainable in the laboratory, although they are excited to luminosity in the hotter stars. Any gas may be made luminous, however, by the passage of an electric discharge through it. In some cases, such as the mercury arc, the temperature is high, while in other cases the light is emitted at a very low average temperature of the source. As examples may be mentioned the various types of phosphorescence, some of which are most active at temperatures as low as that of liquid air, the aurora due to electrical discharges through the highly rarefied and very cold upper atmosphere, and the light emitted by fire-flies and glow-worms.

605. Rectilinear Propagation.—One of the earliest observations concerning light was that, in a homogeneous medium, it travels in straight lines. These lines of propagation or "rays" may be made to alter their direction by two methods—by reflection, when they fall on the boundary between two media, or by refraction, when they pass obliquely from one medium to another, or through a medium of varying density.

606. Shadows and Eclipses.—Rays pass in straight lines by the edges of an obstacle, so that the space behind it is screened

from the light. When the light comes from a very small or "point" source the shadow would be sharply defined if the propagation were strictly rectilinear; as a matter of fact, close observation shows in all cases that the light fades gradually into the shadow. This very significant fact proves that light travels only approximately in straight lines; there is always more or less lateral spreading. Strictly speaking, there is, then, no such thing as a ray of light, if we mean by this term propagation along a geometrical line. The explanation of this spreading will be given later (§676 *et seq.*).

A more obvious cause of the lack of sharpness in shadows is to be found in the fact that most sources of light are not even approximately points, but are of finite area. This gives rise to the distribution of light and shadow shown in Fig. 499 and Fig. 500. The first represents the shadow cast by an object larger

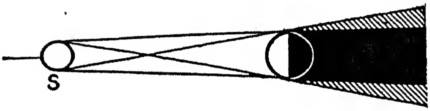


FIG. 499.

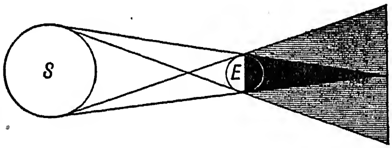


FIG. 500.

than the source; the second, that due to an object smaller than the source, for example the shadow of the earth due to the sun. In each case there is a region of complete shadow behind the obstacle, called the umbra, into which no light from any part of the source can enter. Around this there is a region called the penumbra, which receives light from a part of the source, the effective portion of the latter increasing in going outward from the umbra. When the moon lies entirely within the shadow cone of the earth it is said to be totally eclipsed; when only a portion of the moon passes through the umbra it is partially eclipsed.

607. Parallax.—This well-known phenomenon depends upon the rectilinear propagation of light. By parallax is meant the apparent displacement of an object due to the real displacement of the observer. For example, if the observer moves from O_1 to O_2 (Fig. 501) A will appear to be displaced an angular distance $\alpha + \beta$ to the left with reference to B. That object which seems to be displaced in a direction opposite to the motion of the observer is evidently the nearer. To one traveling on a railroad train

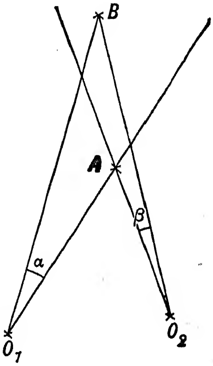


FIG. 501.

objects near at hand appear to be moving backward, those at a distance in the same direction as the observer. If two objects are coincident in position or equally distant their relative parallax vanishes. This gives a useful method of finding the apparent position of the image formed by a lens or mirror, or of focusing the cross thread of a telescope. When the latter and the image of a distant object are both distinctly seen and have no relative parallax, they are coincident in position and both in focus.

In astronomy horizontal parallax is defined as the angle subtended by the semi-diameter of the earth from any body of the solar system. Annual parallax is the angle subtended by the semi-diameter of the earth's orbit from the fixed stars. The distance between the sun and the earth may be determined by observing the transit of an inferior planet, Venus for example, across the sun's disk. Observers at A and B (Fig. 502) note the instants at which Venus appears to enter the sun's disk as viewed from their respective stations. From the interval between these two contacts and the known angular velocity of Venus around the sun the angle α may be determined, and from that and the base line AB the distance of the sun may be calculated. Of course correction must be made for the motion of the earth between the instants of contact.

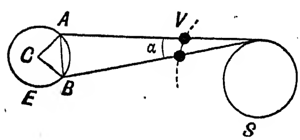


FIG. 502.

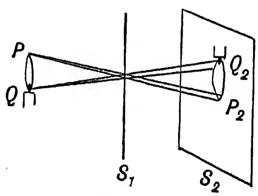


FIG. 503.

608. Pinhole Image.—Another effect of the approximately rectilinear propagation of light is the formation of an inverted image of the source by light passing through a small orifice such as a pinhole. If any source, for example a candle, is placed opposite such a hole in a screen S_1 (Fig. 503) light from the point P will pass through the opening in a narrow cone or *pencil* and illuminate a small patch at P_2 on a screen S_2. Light from Q will form a small patch at Q_2, and light from any other point of the flame will fall on a corresponding point of the screen S_2. The group of patches will in form, color, and relative brightness reproduce the candle flame, but evidently inverted in position. The pinhole forms an image like that due to a condensing lens, but the total light in the pinhole image will be less than that formed by the lens in the proportion of the area of the pinhole to that

of the lens. As the image is due to a group of overlapping patches, it will not be so sharp in outline as that made by the lens. The blurring will increase when the size of the opening is increased or when the source is brought near the screen, thus increasing the angle of the transmitted cone. The object and its image subtend equal angles at the pinhole, so that their linear magnitudes are in the same ratio as their respective distances u and v from the screen S_1. This is true also of images formed by any other optical device, such as a mirror or lens. Landscape photographs of great softness and beauty may be made by the use of the pinhole camera.

609. Reflection, Regular, Diffuse.—When light falls on a smooth polished surface it is reflected in a definite direction. This is called regular reflection. The plane including the direction of the incident light and the normal to the surface at the point of incidence is called the *plane of incidence.* The angle between the incident pencil and the normal to the surface is called the *angle of incidence;* that between the reflected pencil and the normal is called the *angle of reflection.* Experiment shows that: (1) *the angle of reflection is equal to the angle of incidence;* (2) *the reflected pencil lies in the plane of incidence.* It is evident from the first law that if a mirror is rotated through a given angle about an axis perpendicular to the plane of incidence, the reflected pencil will be rotated through twice that angle.

When light falls on a rough unpolished surface it is reflected in all directions. This is called diffuse or irregular reflection. There is no essential difference between regular and diffuse reflection, except that in the latter we may imagine reflection to take place from a large number of very small plane surfaces orientated in all directions.

610. Visibility of Objects.—On a clear night, when there is no moonlight, the stars and planets appear against a background of black sky. The space around the earth's shadow cone is filled with sunlight, but we do not see it unless it is reflected from some planet or the moon. If a beam of light is passed through a vessel of distilled water its path is invisible. If a beam of sunlight enters a dark room it cannot be seen unless dust particles are floating in the air. A drop of milk in the water or a little dust stirred up in the room will cause the path of the light to flash out brilliantly. Such experiments show, as might be expected, that light does not excite the sensation of luminosity unless it

enters the eye directly from the source or by reflection. Ordinary objects are visible because they reflect light diffusely into the eye, and they may be regarded as secondary sources of radiation. A perfect reflector would itself be invisible, all the light reflected from it appearing to come from the image of the source, not from the reflector.

611. Transmission and Absorption.—Light travels through some media, for example most gases, glass and water, with scarcely any appreciable diminution of intensity. Other media may transmit little or none, or certain colors only; such media are said to show *general* or *selective* absorption. In cases where absorption occurs there appears to be a loss of radiant energy, but it may be shown that there has been a change to other forms, usually heat (§330 *et seq.*).

Any substance which transmits a large fraction of the incident light without scattering it is said to be *transparent.* As indicated by this term, objects may be seen clearly through such substances. Objects which absorb all the unreflected incident light are said to be *opaque,* and act as perfect screens. Evidently any perfect reflector must also be perfectly opaque, and in general the reflecting power increases with the absorbing power. Substances differ widely in these properties, varying from almost perfect transparency to almost perfect opacity. The most transparent media known show some absorption, which increases with the length of path; hence any substance will become opaque if a sufficient thickness is taken.

No light penetrates to great depths in the ocean, although a layer of water of considerable thickness is transparent. On the other hand, light will penetrate to a slight depth in any medium, so that thin layers of metal or of carbon are found to be transparent. Some substances are selectively transparent; red glass will freely transmit red light, but not the other colors, and a thin sheet of hard rubber, which appears to be opaque, will transmit radiations lying a little outside the red of the spectrum. In fact, when the entire range of radiation is considered, *all* substances are selectively transparent (§711).

Some substances transmit light, but scatter it so that objects cannot be seen clearly through them. These substances are called *translucent.* The effect is caused by diffuse reflection within the medium, due to discontinuity or non-homogeneity of structure,

as in the case of powdered glass, paper, or water containing finely divided particles. Some substances, such as paraffin, are homogeneous and transparent when in the fluid state, and translucent when in the solid state. The latter effect is apparently due to granulation or crystallization.

Observations made in a dark room when the eye is in a sensitive condition show that a small quantity of light is scattered by the molecules of a filtered gas or of a transparent liquid or solid. Most of the light of the sky is due to molecular scattering, the remainder being due to small particles of suspended matter.

612. Refraction.—When light passes obliquely from one transparent medium to another a part is usually reflected, while that which enters the second medium changes its direction abruptly at the boundary. Generally (but not always) in passing from a lighter medium to a denser, the light is deflected toward the normal to the boundary. This is called refraction. Since objects appear to be in the direction from which the light comes, refraction, by changing the course of the light, causes an apparent displacement of the source. An example is found in the classic experiment of Cleomedes, who showed that a coin placed in the bottom of a vessel so that it is barely concealed by the sides of the latter, is apparently lifted into view when the vessel is filled with water (Fig. 504). The object at A then seems to be at A', a point above and somewhat to the right of A, (§645). Similarly, a meter rod dipped obliquely into water appears to be bent, and the divisions seem to be shortened. A pond seems shallower than it is, even when viewed normally to the surface. This change in the apparent distance of objects seen normally through a refractive medium is to be considered as an example of refraction, although there is no deviation of the light. It will be shown in §643 that these effects are the result of differences of velocity of light in the media concerned.

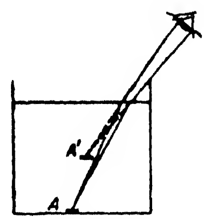


Fig. 504.

613. Intensity of Light.—The intensity of light as estimated by the eye is not capable of precise physical determination. It depends to a large extent upon the color of the light and the sensitiveness of the eye. The only consistent way in which intensity of radiation may be determined or expressed is in terms of energy.

If radiation travels through a homogeneous medium in straight lines, and if the medium is perfectly transparent and does not itself emit radiation, the same total amount of energy must flow per second through any spherical surface concentric with the source. It follows that the *intensity*, or quantity of energy passing through unit area per second, must vary inversely as the square of the distance from the source (§259).

The above conclusion is based upon the assumption that the radiation diverges uniformly in straight lines in all directions. It is not true if the medium is of varying refractivity, on account of partial reflection and of changing divergence of a cone of light in passing from one medium to another. In case a beam is made parallel by a lens or mirror there is no change of intensity with distance except that due to absorption or to imperfect parallelism.

614. Photometry.—The eye can form no exact estimate of the illumination of a surface, but it can determine with great accuracy whether two adjacent surfaces are equally illuminated by light of the same color. Upon this principle are based the different methods of *Photometry* or comparison of light intensities. Two of the simplest and oldest types of photometer are the Rumford shadow and the Bunsen grease spot photometers. In the use of both it is assumed that the light from the two sources compared contains the different colors in the same proportions, making comparison possible.

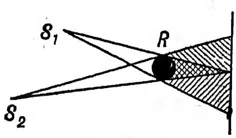


FIG. 505.

In *Rumford's photometer* shadows of a rod R are cast on a white screen by the sources S_1 and S_2 (Fig. 505), one of which is a standard comparison source. By adjusting the positions and distances of S_1 and S_2 the shadows may be made to touch and to be of equal intensity. When this is the case, it is evident that the illumination E from each source is the same at the screen, since each shadow is illuminated solely by the source which casts the other shadow. Let F_1 and F_2 represent respectively the *total luminous flux* radiated uniformly in all directions from the sources S_1 and S_2 (where luminous flux means the energy radiated per second evaluated with reference to visual sensation), and let r_1 and r_2 be the respective distances of the two sources from the screen. Then (compare §259)

$$F_1 = 4\pi r_1^2 E \qquad \text{and} \qquad F_2 = 4\pi r_2^2 E$$

or

$$E = \frac{F_1}{4\pi r_1^2} = \frac{F_2}{4\pi r_2^2}$$

Since there are 4π steradians, or units of solid angle,[1] about any point, $F/4\pi$ will be the luminous flux radiated per steradian. Representing this quantity by I the above relation becomes

$$E = \frac{I_1}{r_1^2} = \frac{I_2}{r_2^2}$$

The quantity I is the *luminous intensity* of the source and is measured in terms of the *international candle.* From the last equations it will be seen that the candle powers or intensities of two sources are directly as the squares of their distances from the screen when the latter is equally illuminated by each. From the above expression, the *illumination* E (which is defined as the luminous flux falling on unit area of surface) produced on a screen by a given source, varies inversely as the square of the distance from the screen to the source.

The *Bunsen photometer* consists essentially of a grease spot on a screen of white paper. Such a spot is more translucent than the clear paper, and for this reason appears darker by reflected light (since there is less light reflected from the spot). If such a screen is placed between sources which equally illuminate it with light of the same quality (same proportions of different colors) the grease spot will disappear. The loss in light reflected from the spot on one side will then be compensated by the increased amount transmitted from the other side. To avoid the effect of any stray light, the sources are interchanged and the readings averaged.

Many other forms of photometers have been devised (§700), but they all employ the principle of equating illuminations as given above.

615. Photometric Units.—France, England and the United States, by international agreement, have adopted an arbitrary standard of luminous intensity which is known as the international candle. This standard is realized in terms of certain specified lamp flames used as primary standards. Two lamps commonly used are the Harcourt pentane lamp, burning pentane vapor, and the Hefner lamp, burning amyl acetate. In each case the lamps must be of specified construction and the flames of specified height. Corrections must be made for the barometric pressure and for the water vapor and carbon dioxide in the atmosphere. The international candle has an intensity equal to 0.1 that of the Harcourt pentane lamp and approximately 1.1 that of the Hef-

[1] If a spherical surface of radius r be drawn with the vertex of a cone as center and th encloses an area S of the spherical surface, the measure of the solid angle of the cone i steradians.

lamp. The most convenient practical standard is an incandescent lamp which has been compared with a primary standard and whose intensity, when a specified current is used, has been certified by the Bureau of Standards at Washington.

The *unit* of luminous flux is the flux which passes out from one international candle in a cone of unit solid angle. This unit is named the *lumen*. Since the illumination is the luminous flux falling on unit area, the *unit* of illumination will depend on the unit of area chosen. The unit most commonly used is one *lumen per square foot* or the foot-candle. It is the illumination produced by one international candle at a distance of one foot. Other units are the *lux* or one lumen per square meter (the illumination produced by one international candle at a distance of one meter) and the *phot* or one lumen per square centimeter.

$$1 \text{ foot-candle} = 10.764 \text{ lux} = 0.001076 \text{ phot}$$

Good illumination for reading should be about 15 foot-candles.

616. Brightness and Lambert's Law.—The *brightness* of a surface emitting luminous energy, such as an incandescent sheet of metal, is defined as the *luminous intensity* (candles) *per unit area of emitting surface normal to the line of sight*. From the previous definition of luminous intensity (§614) it follows that brightness may be defined also as the luminous flux per steradian per unit area of emitting surface normal to the line of sight.

Lambert found, as an experimental fact, that the brightness of a surface is practically independent of the angle from which it is viewed. This evidently must be due to the fact that the luminous intensity decreases with angle in just the proportion that the projected area normal to the line of sight decreases. Thus consider a surface AB (Fig. 506) emitting light. In a direction normal to the surface, let its luminous intensity, in candles, be denoted by I_n. Similarly let its luminous intensity, when viewed at an angle α, be denoted by I_α. Its *brightness*, when viewed normally, is then I_n/AB. When viewed at an angle α, it is $I_\alpha/BC = I_\alpha/AB \cos \alpha$, since BC is the projected area normal to the line of sight. According to Lambert

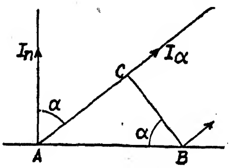


Fig. 506.

$$I_n/AB = I_\alpha/AB \cos \alpha, \text{ or } I_\alpha = I_n \cos \alpha$$

This last relation is known as *Lambert's Cosine* law, and states that the emitted luminous energy per unit angle (candles) from a radiating surface varies as the cosine of the angle between the normal to the surface and the line of sight.

The *practical unit of brightness* is the *lambert*. It is the brightness of a surface having a luminous intensity of $1/\pi$ candle per sq. cm. of projected area. *If* Lambert's law is true for the surface considered, the lambert may be defined more simply as the brightness of a surface radiating one lumen to one whole side (2π solid angle), per sq. cm. of area.

In accordance with Lambert's law, an incandescent sphere when viewed from a distance appears to be a uniformly illuminated disk.

The law does not apply to a surface bounded by an absorbing atmosphere, which will of course exercise greater total absorption in an oblique than in a

normal direction. The sun, for example, which is surrounded by an absorbing atmosphere of gases, appears (as clearly shown in photographs) to be darker at the edges than at the center.

VELOCITY OF LIGHT

617. Velocity of Light.—The sensation of light is produced by a disturbance originating in distant bodies, and it may naturally be assumed that this disturbance travels with a finite velocity. Galileo, about 1600, appears to have been the first to attempt to measure this velocity. His method was substantially the same as that ordinarily used to determine the velocity of sound in the atmosphere.

Two observers stationed at some distance from each other endeavored to note the instants at which flashes of light from one station were observed at the other. The failure of such attempts made it clear that the velocity of light is so great that the time required to pass over ordinary distances is too small to be measured except by methods much more refined than those at that time available. It was natural, therefore, that the first results should have been obtained by astronomical methods, in which the distances employed are those between heavenly bodies.

In 1675 **Römer**, a Danish astronomer, observed that the eclipses of Jupiter's satellites by that planet recur at regularly increasing or decreasing intervals, according to the earth's position with respect to Jupiter. If the first observations are made when Jupiter and the earth are on the same side of the sun and in line with it, the interval between the first and the second eclipse of one satellite is about a day and 18.5 hours, but as the earth proceeds in its orbit the interval between eclipses slowly increases, so that when the earth is on the opposite side of the sun from Jupiter, the eclipse occurs about 16 minutes later than the time calculated from the first observed interval.

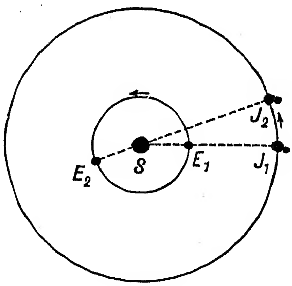


FIG. 507.

Römer explained this as being due to the finite velocity of light. The distance between the earth and Jupiter having in the interval increased by the diameter of the earth's orbit, the last installment of light that comes from the satellite before eclipse has this additional distance to travel and in consequence reaches the earth later by 16 m. 41.6s (according to modern observations). This and the

best determinations of the diameter of the earth's orbit give 298,300 kilometers per second as the velocity of light.

618. Bradley's Method.—Römer's explanation was discredited until long after his death, when an entirely different astronomical method confirmed his views. In 1727 Bradley, later the astronomer royal of England, discovered an apparent displacement of the fixed stars, the direction of the star's *displacement* agreeing with the direction of the earth's *velocity*, in its orbit around the sun. The maximum angular displacement is the same for all stars. Bradley was for a time greatly perplexed by this phenomenon, but the chance observation of the direction of a wind vane on a boat sailing on the Thames, this direction not being that of the wind, but of the resultant of that of the actual wind and that of the virtual wind due to the motion of the boat, suggested to him that the apparent motion of the light coming from the stars might be the resultant of the actual motion of the light and its virtual backward motion due to the forward motion of the earth. If a raindrop S (Fig. 508) is falling in a vertical tube T which is at the same time moving parallel to itself in a direction at right angles to the path of the raindrop, the latter will have a horizontal component of relative motion with respect to the tube and will strike its sides. When the tube reaches T_1, the raindrop reaches S_1 and the *relative* path of S with respect to T is the dotted line PS_1. Similarly a beam of light which actually moves with a finite velocity parallel to the axis of a telescope tube will strike the side of the latter on account of its displacement due to the motion of the earth. If the apparent angular displacement is α, it is evident that $\tan \alpha = v/V$, where v is the component velocity of the earth at right angles to the line of sight and V the velocity of light. Evidently α is the angle through which the telescope must be tilted if a ray of light from a star is to travel along the axis. This apparent angular displacement of the light from the stars is known as the *aberration of light*. The best determination of α, the aberration constant, is 20.479″, which, combined with the known velocity of the earth in its orbit, gives a value for V of 299,714 kilometers per second.

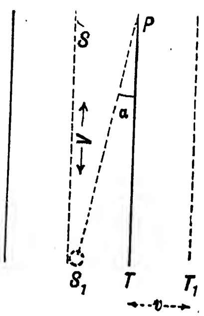


FIG. 508.

619. Fizeau's Method.—The first to make a direct determination of the velocity of light was Fizeau, who in 1849 found the time required for light to pass between Suresnes and Montmartre, near Paris, a distance of 8633 meters. His method was as follows: Light from a source S (Fig. 509) is reflected from a piece of plate glass m, focused by a lens L on the circumference F of a toothed wheel W, and, after passing between the teeth of the wheel, is made parallel by a second lens L_1. From this point the beam travels to the distant lens L_2, which focuses it on a mirror M. From this point the beam retraces its path to the source; but a portion of it will pass through the plate glass m to the eye E, by which it may be observed. If the toothed wheel is rapidly rotated a detached train of light waves will pass through as an opening passes F, travel to M, and return. If in the meantime a tooth has moved into the position F the light will be eclipsed; at twice the speed required for the first eclipse the light will again reach F when an opening is at the point, and will pass to the eyepiece. At three times the original speed of the wheel the second eclipse will occur, and so

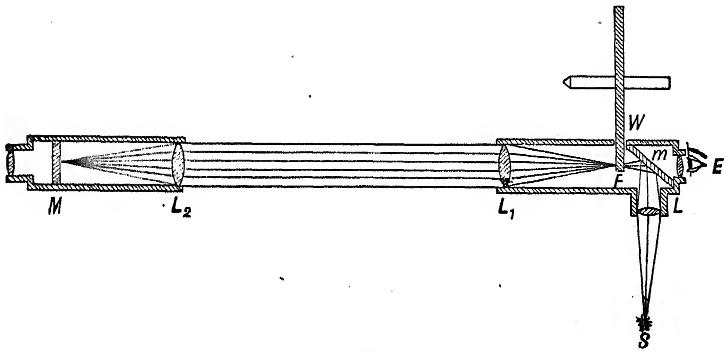


FIG. 509.

on. At speeds permitting transmission of the light the waves will pass and return through the successive openings in intermittent groups, but the light will appear continuous to the eye because of the persistence of vision. From the distance between the wheel and the distant mirror and the rate of revolution of the wheel the velocity of light can be calculated.

The value of V found by Fizeau was 313,300 km./sec. Cornu, using the same method, obtained a mean result of 299,950 km./sec. from several series of experiments.

620. Method of Foucault, Michelson, Newcomb.—In 1862 Foucault determined V by means of the displacement of a beam of light reflected from a revolving mirror. The method was improved by Michelson, who made observations in 1879 at the United States Naval Academy, and in 1882 at Cleveland. Michelson's arrangement is indicated in Fig. 510. Light from a narrow slit S falls on the mirror m and is reflected to a lens L, which throws

it in a parallel beam to the plane mirror M. The beam retraces its path, and if the mirror m is at rest is brought to a focus at S. If, however, m has rotated through the small angle α while the light is passing from m to M and back, the reflected pencil will be rotated through the angle 2α and will form an image of the source at S_1. If the distance between S and $S_1 = d$, that between S and $m = r$, that between m and $M = l$, if n be the number of revolutions of m per second, and T the time required for light to pass from m to M and back,

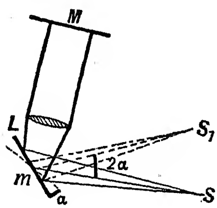


FIG. 510.

$$2\alpha = \frac{d}{r}$$

$$T = \frac{\alpha}{2\pi n}$$

$$V = \frac{2l}{T} = \frac{4\pi l n}{\alpha} = \frac{8\pi l n r}{d}$$

Foucault used a short-focus lens between S and m instead of a long-focus lens between m and M, as in Michelson's arrangement; consequently l was a short distance, not exceeding 20 meters, and the displacement d was only 0.7 mm., even when the mirror revolved 800 times per second. The result obtained by Foucault was 298,000 kilometers per second. In Michelsons's experiments a long-focus lens enabled him to make r large and at the same time to throw a parallel or nearly parallel beam on M, so that the distance l could be increased indefinitely without any considerable loss of light. With a value of $l = 625$ meters, $r = 9$ m., and a speed of 257 revolutions per second, the displacement d was 133 mm.

Newcomb, in 1882, made some further improvements in Foucault's method. The distance l was 3721 meters, between the Washington monument and Fort Myer, in Virginia. The value of V obtained by him was 299,860 km./sec.

Michelson has recently, 1921–26, carried out an extensive investigation at the Mount Wilson Observatory, using a new form of rotating mirror which gives a greatly increased accuracy. The essential change is the use of an octagonal mirror whose speed is adjusted until one face almost exactly replaces an adjacent face, during the time the light is passing from m to M and back (Fig. 510). This distance (Mount Wilson to Mount San Antonio) is 35.4 km. The mirror rotated 528 revolutions per sec. Michelson's final published result is 299,796 $\pm$ 4 km./sec.

621. The Velocity of Light in Different Media, such as water and carbon bisulphide, was determined by Foucault and Fizeau, and also by Michelson, the method of Foucault being used in each case. A long tube filled with the liquid was placed between the mirrors m and M. Michelson found the velocity in air to be 1.33

times greater than that in water, and 1.76 times greater than that in carbon bisulphide. This has an important bearing upon the choice between the emission and the undulatory theories of light (§§623, 643).

The velocity of light from all sources seems to be the same, not being appreciably affected by their intensity. Römer and Bradley used sunlight or starlight, Fizeau and Cornu calcium light, Foucault, Michelson, and Newcomb sunlight, Young and Forbes electric light. In space light travels with the same velocity, regardless of its color. This is shown by the eclipse of a white star by the moon; the star would appear red just before eclipse and blue just after if blue light travels faster than red; but no change of color is observed. It is also shown by the fact that in Michelson's experiment the light was not drawn out in a spectrum. Photographs of the spectrum of the variable star Algol, the light from which has a period of variation of about 69 hours, show that the intensities of the extreme violet and extreme red rise and fall simultaneously, proving that there is no relative retardation between them. In some material media the velocity of light of different colors differs considerably. Michelson found the velocity of blue light in carbon bisulphide to be 1.4 per cent less than that of red. In gases this difference is inappreciable.

Light reaches the earth from the moon in about one second and from the sun in about 8.25 minutes. A small annual parallax has been found in the case of some of the nearer stars, which enables rough estimates of their distances to be made. Light from one of the nearest stars, α Centauri, would require about 3.75 years to reach the earth, and that from Sirius about 17 years. It seems quite possible that a distant star may have been destroyed by an explosion or collision thousands of years ago, and yet be visible to us by light emitted before its destruction and still on its way through space. The changes frequently observed in variable stars must take place years before they are evident to us.

622. The Theory of Relativity.—The *restricted theory of relativity*, proposed by Einstein in 1905, is based on the assumption that the observed velocity of light is not influenced by the relative motion of the source and the observer, nor by motion of the latter with respect to the hypothetical ether. These conclusions were made necessary by the failure of an experiment performed by Michelson and Morley using Michelson's interferometer (§699) in order to detect a drift of the ether with respect to the earth. Mathematically

the assumptions are consistent with the idea that time is a fourth dimension of *space-time*. The theory denies the possibility of experimentally proving the existence of a material ether, and therefore this concept has been largely abandoned. The word "ether" as now used is merely a convenient designation of that property of space by which it transmits light waves.

Einstein later extended his theory to include accelerated motions as well as uniform motions, and this extension is known as the *general theory of relativity*. The predictions of both the restricted and general theories that are subject to experimental test are now considered to be satisfactorily verified. Among these are a variation of the mass of a body with its velocity relative to the observer and the equivalence of mass and energy, both required by the restricted theory. Experiments show that the mass of a body increases with its velocity v according to the relation

$$m = m_0/\sqrt{1 - v^2/V^2}$$

where m is the mass when in motion, and m_0 that when at rest. From this it follows that the kinetic energy E of the moving body is related to the increase of mass $m - m_0$ by the equation

$$E = (m - m_0)V^2$$

In general any quantity of energy E is equivalent to the change of mass $m - m_0$ given by this equation. A striking verification of this is found in recent work on nuclear disintegration (§552). Of the predictions from the general theory, that requiring a deflection of light in passing close to the sun has aroused the greatest interest. The older theories also required such a bending of light by a gravitational field, but by only half the amount calculated by Einstein. The best photographs taken during a total eclipse of the sun show outward displacements of the stars close to the sun which are in excellent agreement with the values from relativity.

THE NATURE OF LIGHT

623. Mode of Transmission.—According to some of the older hypotheses, such as that of Descartes, light is the effect of a pressure instantaneously transmitted through a universal medium. The fact that the disturbance producing light has a finite velocity shows, however, that it is due to motion, not to a static pressure. The

radiation from such bodies as the sun heats substances on which it falls, and may produce chemical changes or electrical effects, which shows that a continuous stream of energy flows from luminous sources. According to our experience, there are only two ways in which energy may be transferred—by the actual projection of material bodies through space or by the transmission of vibrations or pulses through a stationary medium, as illustrated by different types of wave motion. Consequently there have been two rival theories regarding the propagation of light, the *emission* theory and the *undulatory* or *wave* theory. Both theories had been suggested vaguely by ancient Greek philosophers, but their scientific development came at a much later date.

624. Emission Theory.—Sir Issac Newton believed that light is due to the emission of luminous particles ("corpuscles") from the source. He appears to have adopted this hypothesis chiefly because it explained the rectilinear propagation of light for which the wave theory seemed inadequate. Newton showed by prismatic analysis that white light is a combination of many different colors. He attributed difference of color to difference in size of the corpuscles exciting luminosity.

Newton observed that water waves pass around obstacles without sensible disturbance, casting no shadows, and that sound shadows arise only under exceptional circumstances. Reasoning by analogy he could not see why light, if due to wave motion, should not travel around corners instead of in straight lines. He noticed, however, that sound waves had a greater tendency than water waves to cast shadows, and if he had carefully observed the behavior of small waves, such as ripples on water, his objections to the wave theory would probably have been removed. While large water waves pass around a pile or other comparatively small obstacle, ripples are effectually stopped, passing the object on each side without reuniting; there is a well defined region of no disturbance, or shadow. Similarly, sounds of high pitch, due to very short waves, cast well-defined shadows.

The emission theory satisfactorily explains reflection if we suppose the corpuscles to behave like elastic spheres. If such a sphere strikes a reflecting surface at an angle i with the normal (Fig. 511) the tangential component v of its velocity will not be changed. If the magnitude of the reflected component u is unaltered, it follows that the angle r of reflection is equal to the angle i of incidence.

Refraction is also explained if we assume that matter attracts these particles. They will then be subject to a normal acceleration as they approach the boundary, while the tangential component of velocity is unchanged (Fig. 512). If the medium offers no resistance to the motion of the corpuscles (that is, if it is transparent) it follows that the increased velocity should be maintained after entering the second medium, and that the velocity of light should be greater in media more refractive than air. Experiments show that the opposite is true in all cases tested (§621). This is one grave objection to the emission theory. Another is the fact that only a *portion* of the light is refracted, the rest being reflected. Furthermore, if matter attracts light corpuscles, it would be difficult to account for the enormous expulsive forces which would be required to project the particles from luminous sources. We should also expect the speed of the particles to vary with the nature and activity of the source; and yet the velocity of light from a candle appears to be the same as that from the sun.

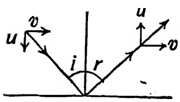


Fig. 511.

Very recently a corpuscular theory of light has been revived (§757) but this new theory is quite different from that of Newton. In particular it predicts a velocity of the light corpuscles in ordinary matter which is in agreement with the experimental facts, whereas Newton's theory, as just noted, is in direct contradiction to the facts.

625. Wave Theory of Light.—Huygens distinctly formulated this theory about 1678. He believed that space is filled with a rare medium, the ether, through which the waves are propagated from luminous bodies. This theory accounts without any difficulty for the ordinary phenomena of reflection and refraction, but was not acceptable to Newton for the reason above stated. For more than a century after Newton's time little progress was made in the subject of light, until in 1802, Thomas Young published a paper "On the Theory of Light and Colors." In this he discussed optical phenomena from the standpoint of the wave theory, and first called attention to the fact, overlooked by Huygens and other advocates of the wave theory, that the effect at any point of space through which light waves are passing is the resultant of the effects of a number of coincident individual waves. The magnitude of this resultant depends not only on the amplitudes, but also on the relative phases of the component waves. If two waves of equal amplitude and moving in the same direction are in the same phase, the displacement at any point is the sum of the individual displacements, and the energy, which is proportional to the square of the amplitude, is four times as great as in a single wave. If the waves

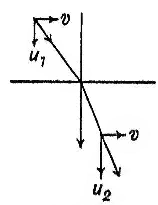


Fig. 512.

are opposite in phase, the resultant amplitude and energy at any point are zero. This effect Young called the *interference* of light waves.

Young devised a simple experiment which may be regarded as a crucial test of the wave theory. Light diverging from the slit S (Fig. 513), which acts as a primary source, passes through two narrow slits S_1 and S_2 very close together and equidistant from S. These act as secondary sources. If a screen be placed beyond these slits a series of colored and dark bands parallel to the slits will be observed on it. If one of the slits is covered the bands disappear. This shows that they are the resultant effect of two superimposed pencils of light alternately re-enforcing and destructively interfering with each other. This is analogous to the interference of mercury ripples described in §258.

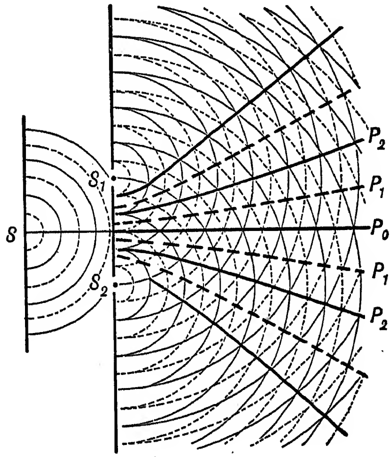


Fig. 513.

It is easy to repeat Young's experiment by ruling two narrow slits very close together on a developed photographic plate and looking through these slits at a distant electric light. The explanation is as follows: Through the slits S, S_1, and S_2 the wave disturbance propagates itself in all directions beyond the respective screens in semi-cylindrical waves having these slits as axes, as may be seen by holding a white screen in front of such a narrow slit on which light falls. It will be seen that the transmitted light diverges very considerably from the axis of the pencil, the amount of divergence increasing as the slit is narrowed (§682). There are, consequently, when two slits are used, two sets of semi-cylindrical light waves diverging from these slits and crossing each other, as shown in Fig. 513. Along SP_0, every point of which is equidistant from S_1 and S_2, waves of all lengths from the two sources will always meet in the same phase, and there will be a maximum of white light on the screen at P_0. Along the dotted line ending at P_1 the distances of any given point from the two sources differ by half a wave-length;

there is destructive interference along this line and a minimum for the corresponding color at P_1. Along the line ending at P_2 the difference between the distances of any given point from the sources is a whole wave-length, so that along this line waves of the same length meet in the same phase and there is a maximum for the corresponding color at P_2. At any point P_n for which $S_1P_n - S_2P_n = N\lambda$ (N being any whole number) there will be a maximum; where $S_1P_n - S_2P_n = (N + \frac{1}{2})\lambda$ there will be a minimum.

Let P_2B be equal to P_2S_2 (Fig. 514). Then $S_1B = \lambda$. Denote S_1S_2 by a. Since a is very small, S_2B is very nearly perpendicular to S_1P_2. Hence

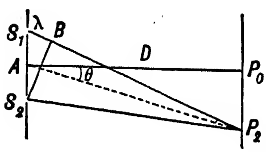


FIG. 514.

$$\lambda = a \sin \theta$$

From this λ can be deduced. Measurements show that it is very small, being about 0.000065 cm. for red light and 0.00004 cm. for violet light.

If AP_0 be denoted by D, $\sin \theta = P_0P_2/D$, since AP_0 and AP_2 are very nearly equal. Hence $P_0P_2 = D\lambda/a$. Thus, for a given distance of the screen, the width of a band varies directly as the wave-length and inversely as the distance between the slits.

626. Relation between Velocity, Wave-length, and Frequency.—If white light falls on the slits the interference bands in general disappear, except in the neighborhood of P_0. The central band at P_0 is white, while the inner edge of the first band on either side is violet, and the outer edge red. Beyond that there is a general overlapping of colors. This shows that the wave-length is different for light of different colors, and that the wave-length of violet light is less than that of red. The central band is of course white, as all colors have a maximum at this point, regardless of their wave-length. From the relation $n\lambda = V$ (§246) where n is the frequency of vibration, λ the wave-length, and V the velocity of light, it is evident that when V changes either n or λ or both must change. If Young's experiment be performed in a medium such as water, it is found that the width of the bands in water is to their width in air as the velocity of light in water is to that in air. Hence $\lambda_1/\lambda = V_1/V$, and n is constant.

627. Huygens' Principle.—Huygens assumed that a wave is propagated by every point of the medium in a wave front acting as a new center of disturbance, as has already been explained and illus-

trated in the case of water waves (§256 and Fig. 166*f*). The resulting wave front is the enveloping tangent plane to the wavelets starting from these centers, as shown in Fig. 515.

The points *a*, *b*, *c*, etc., between *A* and *B* (Fig. 515), taken as close together as we please, act as centers of disturbance. Along the tangent plane *A′B′* the different waves are all in the same phase, and each point in this new tangent plane becomes a new center of disturbance, so that the resultant wave travels forward as rapidly as the disturbance is propagated from point to point of the medium.

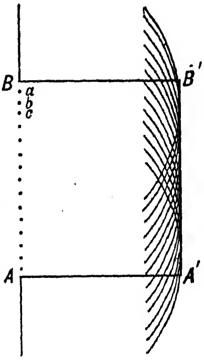


FIG. 515.

The waves move forward without hindrance, because there is no existing displacement to oppose them; they do not travel backward, because there is a force due to the existing displacement on the side from which the waves come sufficient to nullify the backward component of the displacement due to each successive center of disturbance. It is like the propagation of a shove through a line of people, or of elastic spheres of the same mass and elasticity; that in front is not braced to withstand the impulse, while the reaction on the one communicating the impact is expended in overcoming its forward momentum.

628. Origin and Properties of Light Waves.—Sources of light are usually bodies at high temperatures. According to the mechanical theory of heat, high temperature corresponds to a violent agitation or vibration of the ultimate particles (molecules, atoms or electrons) (§151) of matter. We may imagine that the displacements of these charged particles send out electric waves through space in much the same way that a tuning fork generates sound vibrations in air.

So far we have presented no evidence to show whether these waves are longitudinal, like those of sound, transverse, like those in a stretched wire, or of a more complex character, like water waves. In §734 it will be shown that the displacements in these waves must be transverse to the direction of propagation.

We may now, as a working hypothesis, assume that light is due to transverse periodic electric displacements in a universal medium, set up by the agitation of the particles of matter; that these waves are of different lengths (periods of vibration), but are all very short; that different colors correspond to different rates of vibration; and that waves of all lengths travel with the same velocity in free space,

but with different velocities in matter. All experimental facts are in harmony with these assumptions. For a discussion of recent theories of the nature of light see §757.

In the following pages the word *ray* will often be used as a matter of convenience, *meaning thereby merely a normal to the wave front, which indicates the direction in which the wave is moving at the point considered.* The definition applies only to isotropic media (§§163, 742).

REFLECTION

629. Reflection from a Plane Surface.—A wave diverging from the source S (Fig. 516) falls on a plane mirror MN. If the mirror were absent, the wave would at a given instant occupy the position $AMPNB$. With the mirror in place, each element of the original wave when it reaches the mirror becomes the center of a reflected wavelet, just as it would have contributed a wavelet from the same point to form the resultant wave MPN if the mirror were absent. If, therefore, a number of circles tangent to MPN be described about centers a, b, c, etc., on MCN they must touch both the imaginary wave MPN and the reflected wave MQN. The arcs MPN and MQN are evidently similar and equal and have equal radii of curvature. If S_1 is the center of curvature of the reflected wave, $SC = CS_1$, the line SS_1 is normal to the mirror, and S_1 is as far behind the latter as S is in front of it. If the eye is at E, any point reached by the reflected wave, the pencil of light entering the pupil will be focused on the retina. As the vertex of this cone is virtually at S_1, the image of the source will appear to be at that point. From the diagram it is evident that the angles i and r are equal, that is, **the angles of incidence and of reflection are equal.**

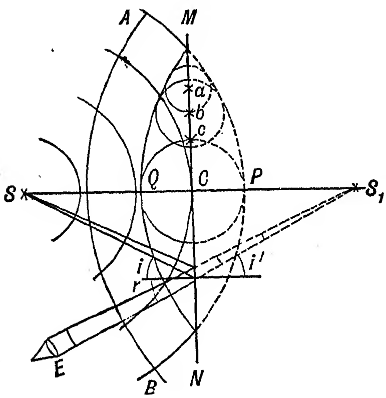


FIG. 516.

630. Focus.—The source or center of curvature of a family of waves, either divergent or convergent, is called a *focus*—literally a hearth or source of radiation. The point S from which the waves actually come is called a *real focus;* the point S_1 from which they appear to come is called a *virtual focus.* The points S and S_1 are

conjugate foci. Since the conjugate focal distances in the case of a plane mirror are equal, it is evident that if the mirror be displaced a given distance parallel to itself the image will be displaced twice that distance.

631. Images.—If $A'B'$ is the image of AB (Fig. 517), it may be shown as above that the image of each point is as far behind the mirror as the point itself is in front, and on the same normal; and that, consequently, the image and the object are symmetrically placed with respect to the mirror and are of the same size.

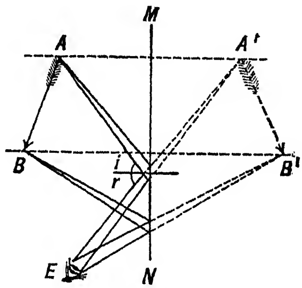


FIG. 517.

632. Multiple Reflection.—Fig. 518 shows how these images are situated in the case of multiple internal reflection from surfaces AB and CD parallel to each other. The position of these images is readily determined by the fact that the image of the first order in each surface is as far behind the surface as the source is in front, and on the same normal to the surface. The two images of the second order are fixed in the same way, by considering the images of the first order to be the sources, and so on *ad infinitum.* It is easy to see that when the mirrors AB and CD are inclined at an angle α (Fig. 519) there are

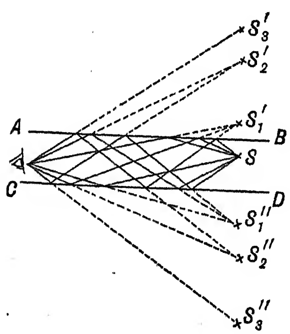


FIG. 518.

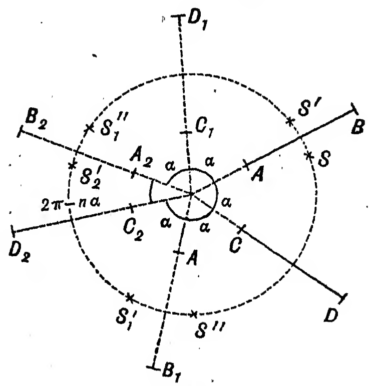


FIG. 519.

multiple images of the mirrors as indicated by the dotted lines, and that the successive images of an object S are symmetrically placed on each side of each mirror image and located in a circle about the point of intersection of the mirrors.

633. Reflection from Curved Surfaces.—If a wave is reflected from a curved surface the curvature of the reflected wave is changed,

unless it exactly conforms to the mirror surface at incidence. Experience shows that only in a few cases is the reflected wave spherical or approximately so, and only in such cases can a definite image be formed. The ordinary type of curved mirror is that with a spherical surface. The reflected waves are approximately spherical if the diameter of the mirror is small compared with its radius of curvature. In order to determine the position of the center of curvature and the conjugate focal relations for spherical mirrors the following very simple mathematical relation is required.

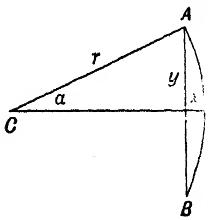


FIG. 520.

634. Relation between Radius of Curvature and Sagitta of Arc.—Consider the arc AB, with center of curvature C, and radius r (Fig. 520). The distance x on the bisecting radius of the arc included between the arc and the chord $AB = 2y$ is called the sagitta of the arc. To determine the relation between r and x write

$$r^2 = y^2 + (r - x)^2 = y^2 + r^2 - 2rx + x^2$$

Therefore,

$$2rx - x^2 = y^2$$

$$x = \frac{y^2}{2r - x} = \frac{y^2}{r(1 + \cos \alpha)}$$

It is found that if the angle α is very small, not more than two or three degrees, the mirror will give a well-defined image. If the angular aperture 2α of the mirror is greater than four or five degrees spherical aberration becomes noticeable (§640). For all mirrors which give satisfactory images x may be neglected in comparison with r, or $\cos \alpha$ regarded as equal to unity, so that within the limits of errors of measurement

$$x = \frac{y^2}{2r}$$

635. Concave Mirror.—The source is at a distance u from a concave mirror MN (Fig. 521) with center of curvature at C and radius r. The waves incident on the mirror have a radius of curvature u, with a sagitta AB. Reflection begins at M and N while the vertex of the wave has still to travel the distance BD before reflection begins at D. When the vertex reaches the mirror the edges of the wave have travelled a distance $BD = AD - AB$ along

MS_1 and NS_1. If the reflected wave is spherical it must have a definite center of curvature S_1 and radius v, with sagitta DE. At the instant when reflection begins at D the incident wave, the mirror, and the reflected wave have a common point of tangency at D. If the angular aperture of the mirror is so small that the cosines of the angles α, β, and γ may be considered as equal to unity (these angles are exaggerated in the figure for the sake of clearness) we may consider the portions of the wave reflected from M and N to move parallel to the axis rather than in the directions MS_1 and NS_1. Hence

$$ED - AD = AD - AB$$
$$AB + ED = 2AD$$

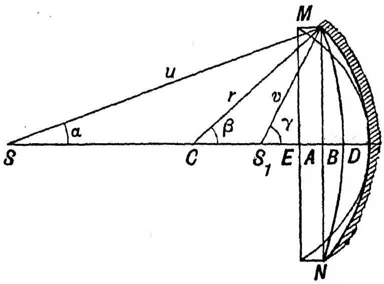


FIG. 521.

It is not convenient to measure sagittae, but by using the relation developed in §634 the above expression can be transformed into one involving only the easily measured distances r, the radius of curvature of the mirror, u, that of the incident wave, and v, that of the reflected wave. The semichord y has the same value for all the arcs concerned, so that the common factor $y^2/2$ may be cancelled when $y^2/2r$ is substituted for AD, with similar substitutions for AB and ED. The final result is

$$\frac{1}{u} + \frac{1}{v} = \frac{2}{r}$$

The justification for the somewhat inexact assumptions made in deriving this formula is found in the fact that it agrees with experimental observations within the limits of error of measurement, provided α, β, and γ are small angles.

A beam of light is always reversible in direction; hence, if the source is at S_1, the image will be at S.

If the source is at a great distance from the mirror the incident wave is practically plane (parallel beam), and u is infinite. The corresponding value of v is called the *principal focal distance* f. S_1 is then the *principal focus*, and the above equation becomes

$$\frac{1}{\infty} + \frac{1}{f} = \frac{2}{r} \text{ or } f = \frac{r}{2}$$

Hence the principal focal point is half way between the mirror and its center of curvature. The conjugate focal relation may now be written,

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$$

If $u > r$, $v < r$. The image is between C and the mirror.
If $u = r$, $v = r$. The image is at the center of curvature C.
If $u < r$ but $> f$, $v > r$. The image is beyond C.
If $u = f$, $v = \infty$. The reflected light is parallel.

If $u < f$, v is a negative quantity. Fig. 522, which illustrates this case, shows that the center of curvature of the reflected wave is behind the mirror. It is a *virtual focus,* since the waves do not actually diverge from that point.

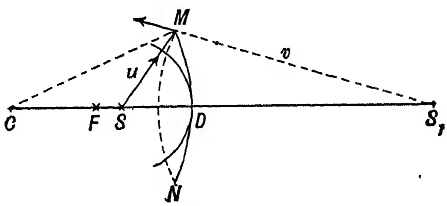


Fig. 522.

Writers differ in their conventions regarding the signs of conjugate focal distances and radii of curvature. The rule most easily remembered and applied seems to be the following:

Consider each of the quantities u, v, r as positive when it denotes a distance measured in front of the mirror.

From this rule it is evident that a positive value of v or f indicates a real focus, a negative value a virtual focus.

636. Convex Mirror.—Proceeding as in the previous case, if MEN is a convex mirror, DE the sagitta of the incident wave, and FG the sagitta of the reflected wave and v its radius (Fig. 523),

$$DE + EF = GF - EF$$
$$DE - GF = -2EF$$
$$\frac{1}{u} - \frac{1}{v} = -\frac{2}{r} = -\frac{1}{f}$$

where, provisionally, v, r, and f, may be considered as mere magnitudes affected with negative signs in the formula.

Comparing this expression with that deduced for a concave mirror, we see that it will become identically the same if we agree to consider v, r, and f as negative in accordance with the rule just given.

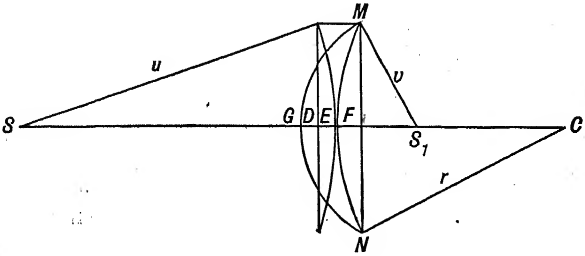


FIG. 523.

The general formula applicable to all mirrors is, therefore,

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$$

u being positive for any real object, and f positive for a concave, negative for a convex mirror. When $f = \infty$ we have the case of a plane mirror.

If we make f negative, as is the case for a convex mirror, we see that v is always less than f and negative. If, however, the light incident on the mirror is convergent to a point at a distance $-u$ behind the mirror, v may become positive: so that a convex mirror may give a real image of a virtual object.

637. Geometrical Method.—The same results may be obtained by applying the law of plane reflection to "rays" without any hypothesis as to the nature of light. The ray SD (Fig. 524) will, as it is incident normally at D, be reflected back on itself. The ray SP will be reflected at P so that the angles i and r are equal.

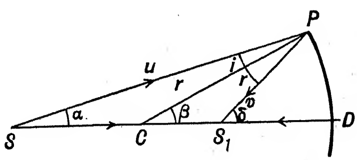


FIG. 524.

The intersection of these two reflected rays will fix the position of the image S_1. From a well-known geometrical relation we have

$$\frac{SC}{CS_1} = \frac{SP}{PS_1}$$

Therefore,

$$\frac{u - r}{r - v} = \frac{u}{v}$$

from which

$$vr + ur = 2uv$$

Dividing through by uvr,

$$\frac{1}{u} + \frac{1}{v} = \frac{2}{r}$$

The assumptions made in this case are that $SD = u$ and $S_1D = v$, which is a sufficient approximation to the truth when the angles α, β, and δ are small.

The formula for the conjugate focal relations of a convex mirror may be derived in the same way.

In some cases it is more convenient to use the geometrical or ray method than that of waves; but it must always be remembered that these "rays" merely represent normals to the wave front.

638. Images Formed by Spherical Mirrors.—If any two rays be drawn from any point of a source, the point of their intersection

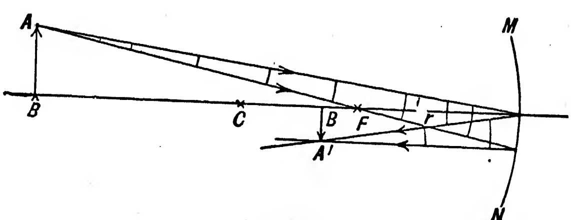


FIG. 525.

after reflection will fix the position of the corresponding point of the image. Any pair of rays will do, but for convenience two of the following are usually chosen, because their course after reflection is easily determined: the ray parallel to the axis, which after reflection passes through the principal focus; that which passes through the center of curvature and returns on itself; that which passes through the principal focus, which becomes parallel to the axis after reflection; that which is incident at the intersection of the mirror with its axis and is reflected making $\angle r = \angle i$. The last two seem the most instructive, the first two are usually the easiest to draw.

The construction of the images formed by a concave and by a convex mirror is illustrated by Figs. 525, 526, 527, where the points A' and B' are located by using the pair of rays last mentioned. (The student should redraw these images, using the other pair of rays.) In the first, the image is real and inverted; in the second and third the images are virtual and erect.

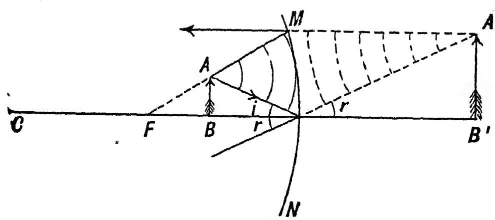


Fig. 526.

639. Magnification.—Since the angle subtended by the object at the mirror is i, while that subtended by the image is the equal angle r, it is evident that the relative sizes of the object and image are to each other as their respective distances from the mirror.

$$\therefore\ o/i = u/v$$

The real image formed by a concave mirror may be of the same size as the object, or larger, or smaller; the virtual image is always

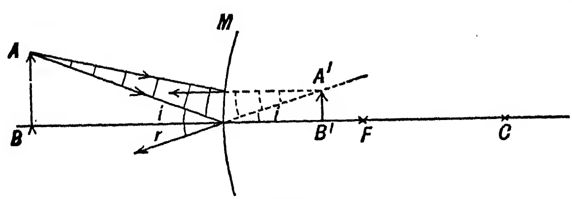


Fig. 527.

larger, since $v > u$. The virtual image formed by a convex mirror is always smaller than the object.

640. Spherical Aberration and Caustic Curves.—If a converging wave is truly spherical there is a perfect focus at its center of curvature. As a matter of fact, the waves reflected from a spherical mirror are not perfectly spherical, except in the special case where the source is at the center of curvature of the mirror. The reflected rays (normals of the reflected waves) are tangent to the curve HF (Fig. 528), which is called a *caustic curve*. The cusp F of this curve

corresponds to the focal point of a mirror of small aperture. The light reflected from the sides of a cup containing coffee or milk plainly shows a caustic curve on the surface of the liquid.

The deviation from a spherical shape of waves reflected from a mirror of large aperture is called *spherical aberration.*

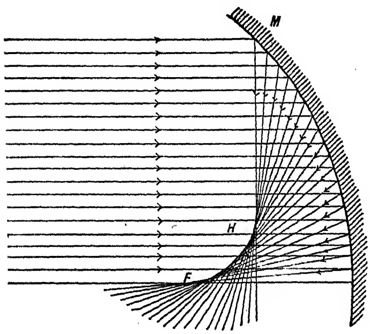


FIG. 528.

If light is obliquely incident on a mirror, the reflected waves are not spherical, even when the aperture is small, but have different radii of curvature in planes at right angles. As a result, a point image of a point source cannot be obtained, but there are two elongated images at right angles to each other and in different positions, which are called *focal lines.*

The origin of the focal lines is clearly seen if we consider the mirror MN (Fig. 529) to be part of a large mirror MNO, on the axis SB of which the source S lies. Constructing the reflected rays incident at different points of this mirror, it is clear that, while the focal cusp of the entire mirror is at S_1, all the rays coming from MN intersect approximately at the point F_1. The diagram gives a cross-section of the incident and reflected rays. If corresponding diagrams be drawn in planes through SB and intersecting the mirror MN,

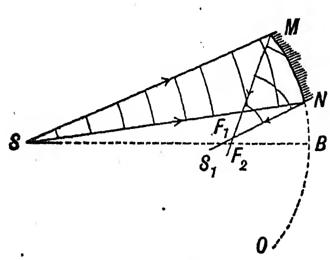


FIG. 529.

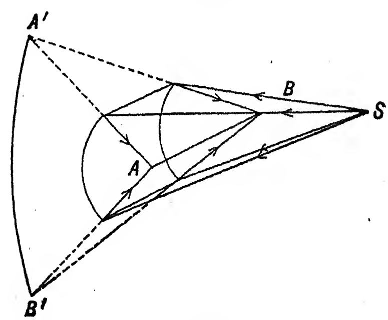


FIG. 530.

the points corresponding to F, will lie on the arc of a circle with its center on the line SB. This is the primary focal line, which will appear on a screen placed at F_1 as a narrow curved strip. After passing F_1 all the rays reflected from MN will intersect the axis SB at various points between S_1 and F_2 (since all the planes of incidence contain SB). A screen placed in the position SB will show a narrow elongated patch of light, S_1F_2, the secondary focal line. If the screen is

at right angles to the reflected pencil the patch of light will be approximately a lemniscate or figure 8.

641. Cylindrical Mirror.—A parallel beam incident on a cylindrical surface is brought to a real or virtual line focus. The image of a point source is likewise a line. Such mirrors and the reflected pencil are said to be *astigmatic.* (A pencil symmetrical about an axis, that is, having a point vertex, and thus giving a point image of a point, is said to be *homocentric.*) In the case of a concave cylindrical mirror, if the point source lies outside the principal focus, there will be a real image AB and a virtual image $A'B'$ in planes at right angles to each other, as illustrated in Fig. 530.

642. Paraboloidal, Ellipsoidal, and Hyperboloidal Mirrors.—The light from a point source at one focus of an ellipsoidal reflector will be brought without aberration to the other focus, a real image being formed. Light from a source at one focus of a hyperboloidal mirror will have a virtual focus at the conjugate focus of the mirror. If the source is at the focus of a paraboloidal mirror, the light will be reflected in a parallel beam; and parallel light will be brought without aberration to a real focus by such a mirror. Paraboloidal mirrors are used for headlights and for the mirror of a reflecting telescope (§693).

REFRACTION AND DISPERSION

643. The ancients were acquainted with the fact that a beam of light is more or less deviated in passing from air to water. The **Law of Refraction** was first discovered in 1621 by *Willebrord Snell.* He found by experiment that **the ratio of the sines of the angles of incidence and of refraction is constant at the boundary between two (isotropic) media.** The ratio $\sin i/\sin r$ (Fig. 531) is called the *index of refraction* and is denoted by n. The angle of incidence is usually measured in air.

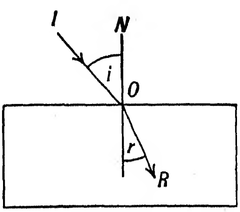


FIG. 531.

It was shown by Huygens that refraction is very simply explained by assuming a change of velocity in passing from one medium to another. Direct measurements by Foucault, Fizeau, and Michelson show that light travels with different velocities in air, water, and carbon bisulphide (§621).

Consider a plane wave AC incident obliquely on the smooth plane surface of separation between air and another transparent medium (Fig. 532), the velocity in air being V_1 and that in the second medium V_2. A spherical wave will diverge from the point A into the second medium when the disturbance reaches that point, and later other spherical waves successively diverge from B' and C'. While the wave travels in the first medium a distance

$CC' = V_1t$, the wave from A will travel the distance $AA' = V_2t$ in the second medium. The disturbance from B will in the same time travel a distance $BB' + B'B'' = (V_1 + V_2)t/2$, if B is half way between A and C. Since $B'B'' = \frac{1}{2}AA'$, a tangent plane can be drawn from C' to the two circles with centers at A and B'. It is easily shown by this method that the waves from all points in the

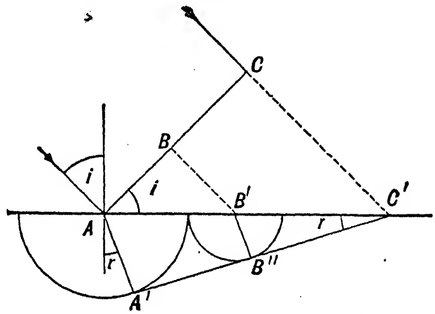


FIG. 532.

original wave front will be tangent to the same plane, the new wave front in the second medium. Further,

$$CC' = AC' \sin i = V_1t$$
$$AA' = AC' \sin r = V_2t$$

Therefore,

$$\frac{\sin i}{\sin r} = \frac{V_1}{V_2} = n$$

The physical significance of the constancy of the sine ratio discovered by Snell thus becomes apparent. The student should always think of the index of refraction as being the ratio of the velocities of light in the two media, rather than as the ratio of the sines of two angles. The latter mode of statement conveys no clear physical idea, and, moreover, seems to break down in the case of normal incidence.

If the measured values of n in the table (p. 579) are compared with the ratio of the velocities determined directly (§621), it will be seen that there is good agreement for water, but not for carbon bisulfide. The discrepancy is due to the fact that it is only the velocity with which energy is transferred (the so-called group velocity) that can be measured in an experiment like that of Michelson, and the group velocity differs from the wave velocity by an amount that depends on the dispersion of the medium (§650). In the case of carbon bisulfide the dispersion is large, thus accounting for the discrepancy.

644. Medium with Parallel Surfaces.—An incident pencil will be deflected in one direction on entering the second medium of thickness t and an equal amount on reëntering the first medium as shown in Fig. 533. The course of the pencil will then be parallel to its original direction, but there will be a lateral displacement AB.

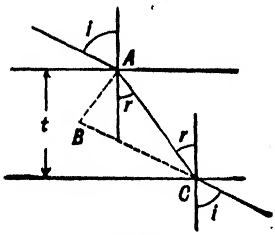


FIG. 533.

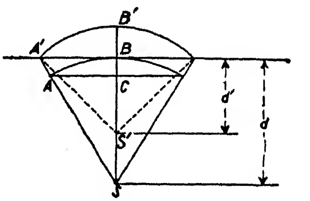


FIG. 534.

645. Image due to Refraction at Plane Surface.—When an object is viewed normal to the boundary (Fig. 534) there is no lateral displacement, but only an apparent change in distance. Waves from an object S at a distance d below the surface of the medium travel with a velocity V_2 to the point B, where the vertex of the wave enters air, in which the velocity is V_1. The disturbance then travels a distance $BB' = V_1t$ in air while another portion of the wave still within the second medium travels the distance $AA' = V_2t$. The center of curvature of the emergent wave is at S', a distance d' below the surface. There is a virtual image of the source at this point. If the cone has only a small divergence, $AA' = BC$, the sagitta of the wave in the refracting medium, BB' is the sagitta of the wave in air, and $d = AS$ and $d' = A'S'$ their respective radii of curvature; hence, from the relation previously used (§634),

$$AA' = y^2/2d = V_2t$$
$$BB' = y^2/2d' = V_1t$$

Therefore,

$$\frac{d}{d'} = \frac{V_1}{V_2} = n, \text{ or } d = nd'$$

The angle of the cone of light entering the eye is limited by the size of the pupil, and is, therefore, very small, so that the use of the above method is justified. The apparent depth of the object below the surface is $d' = d/n$. There is an apparent displacement toward the observer amounting to $(d - d') = (n - 1)d/n$. It is thus made clear why the depth of a pond appears to be less than it actu-

ally is, and why objects immersed in water appear to be shortened vertically. Since the index of refraction is about 1.33, a pond six feet deep seems to be only about four and a half feet in depth.

If the cone is wide there is considerable aberration, as shown in Fig. 535. The image may appear at any point of the caustic curve, depending on the position of the eye. Thus if the eye is at E, the object P is displaced upward and laterally to the position Q (compare Fig. 504).

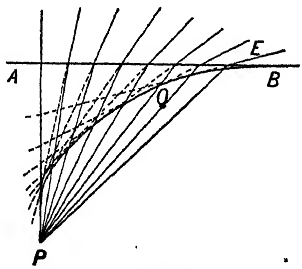


FIG. 535.

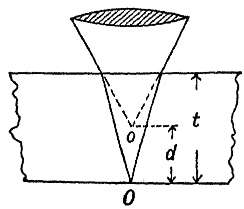


FIG. 536.

The index of refraction of plane parallel plates may be obtained from the relation deduced above. A microscope is focused on a small object on a table, such as a pencil mark O (Fig. 536). When the plate is placed over the mark it will be necessary to raise the microscope a distance d to bring the virtual image o into focus. The apparent depth of the object below the surface is $t' = t/n$, and $d = t - t' = t - t/n$. Hence

$$n = \frac{t}{t - d}$$

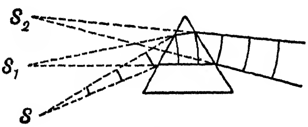


FIG. 537

646. Prism.—If light waves pass through a transparent medium bounded by plane surfaces which are not parallel, the deviation of the incident pencil on entering the first surface is not compensated on emerging from the second surface. If the source is at S (Fig. 537) the image, or center of curvature of the wave within the prism, is at S_1 and that of the emergent wave is at S_2. To determine the deviation of the pencil and the positions of the foci S_1 and S_2 it is convenient to follow the course of a given wave normal or "ray." The intersection of pairs of such rays will fix the position of the desired foci or centers of curvature of the waves.

647. Deviation. Minimum Deviation.—The total deviation of a given ray is $D = D_1 + D_2$ (Fig. 538).

$$D_1 = i_1 - r_1;\ D_2 = i_2 - r_2;$$
$$D = i_1 + i_2 - (r_1 + r_2)$$
But $r_1 + r_2 = A$, since
$$B + A = 180° = B + r_1 + r_2.$$
Therefore, $D = i_1 + i_2 - A$

It is easily shown, experimentally or mathematically, that D has a minimum value when $i_1 = i_2$, in which case the incident and emergent rays are symmetrical with respect to the refracting angle of the prism. In this case

$$i_1 = i_2 = \frac{D + A}{2}$$
$$r_1 = r_2 = A/2$$

Therefore,

$$n = \frac{\sin i}{\sin r} = \frac{\sin \frac{1}{2}(D + A)}{\sin \frac{1}{2}A}$$

This relation is commonly used for determining the index of refraction of substances in the prismatic form. The angles of the prism and of minimum deviation are measured with a spectrometer (§697). As the index is not the same for different colors, it is evident that the prism can be set at the angle of minimum deviation for only one color at a time.

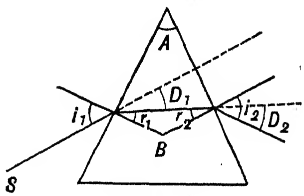


FIG. 538.

648. Dispersion of Color.—The index of refraction of any given substance varies with the color, or wave-length; consequently the deviation caused by a prism will not be the same for all colors. Consider a narrow source S such as an illuminated slit parallel to the edge of the prism (Fig. 539). If the source emits red light alone, a virtual red image of the slit is observed at R. If green and violet light are also emitted, a green and a violet image are seen at G and V. Real images of these colors may be formed at R', G', and V' by a lens. Such a group of line images is called a *line spectrum.* This separation of the colors is called *dispersion.* If the source emits waves of an infinite number of lengths included between the red and the violet, the infinite number of partially

overlapping images of the slit will form a *continuous spectrum.* If the slit is wide the different colors will greatly overlap, and the spectrum is said to be impure. There is less overlapping when the slit is narrowed; but, since no slit can be made infinitesimally narrow, it is manifestly impossible to obtain a perfectly pure spectrum.

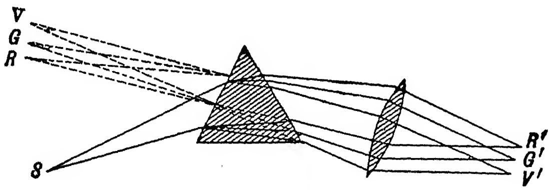


FIG. 539.

649. Fraunhofer Lines.—If a wide slit illuminated by sunlight is used, a continuous spectrum appears, apparently like that given by a candle flame. Such a spectrum was observed by Newton. If, however, the slit is very narrow, it will be seen that a number of fine dark lines parallel to the slit cross the spectrum. These lines were first seen by Wollaston in 1802. He observed a virtual spectrum by looking directly through a prism at an illuminated slit. Fraunhofer, about 1815, by the use of better prisms, and by forming a real image of the spectrum with a lens, was able to find several hundred of these lines, which are now usually referred to as Fraunhofer lines. It is evident from this that the solar spectrum differs from that of a candle in not being absolutely continuous. The dark gaps in the position of different colors show the absence of corresponding images of the slit, and therefore the absence of these colors in the sunlight. In the section on Absorption it will be shown that most of these dark lines are due to the absorption of light of definite wave lengths by vapors in the solar atmosphere (§714).

The Fraunhofer lines may be used as reference points in measuring indices of refraction of prisms for different colors. The more prominent lines were labeled by Fraunhofer with letters of the Roman and Greek alphabets. Some of the more important of them are the *A* line (really a group of fine lines), due to absorption by oxygen in the earth's atmosphere; the neighboring *D* lines, due to sodium vapor in the sun; the *C* and *F* lines, due to hydrogen; the *H* and *K* lines, due to calcium. These lines are shown in the reproduction of the solar spectrum (upper part of Fig. 602).

650. Dispersive Power.—The deviation of a particular color by a prism increases with the index of refraction. The angular separation or dispersion between two colors depends on the difference between their respective indices of refraction. If a prism has a very small refracting angle, the angles of incidence, refraction, and emergence of a given pencil transmitted at the angle of minimum deviation will likewise be small, and the sines of these angles may be considered as equal to the arcs; consequently,

$$n = \frac{\sin \frac{1}{2}(A + D)}{\sin \frac{1}{2}A} = \frac{A + D}{A}$$

Therefore,

$$D = (n - 1)A$$

If D_1 and D_3 are the deviations of two given colors, the Fraunhofer lines C and F, for example, and D_2 that of an intermediate color halfway between C and F, $D_3 - D_1$ is the angular dispersion of the extreme colors and D_2 is the mean deviation of the spectrum of angular width $D_3 - D_1$. The *dispersive power* d of the prism is the ratio of the angular dispersion of the two colors to their mean deviation, or

$$d = \frac{D_3 - D_1}{D_2} = \frac{(n_3 - 1)A - (n_1 - 1)A}{(n_2 - 1)A} = \frac{n_3 - n_1}{n_2 - 1}$$

	n_D	$n_F - n_C$	$\frac{n_F - n_C}{n_D - 1}$
Water	1.3330	0.0060	0.0180
CS_2	1.6303	.0345	.0547
Ether	1.3566	.0052	.0149
Alcohol	1.3597	.0062	.0174
Crown glass	1.5160	.0073	.0141
Light flint glass	1.5718	.0113	.0197
Heavy flint glass	1.7545	.0274	.0363
Very heavy flint glass	1.9625	.0488	.0507
Quartz	1.5442	.0078	.0129
Diamond	2.4173	.0254	.0179
Iodide of silver	2.1816	.1256	.1063
Air (0°C., 760 mm.)	1.00024289	.00000295	.0121
H_2	1.00014294	.00000195	.0136
CO_2	1.00044922	.00000460	.0102

Newton assumed that the ratio of dispersion between two given colors to the mean deviation, or the dispersive power, is the same for all substances, but Dollond, in 1757, showed that this is by no means the case. Two different prisms may have the same value of $n_2 - 1$, but very different values for $n_3 - n_1$, or conversely.

The table shows the values of n_D and the dispersive power between the C and F lines for some substances, the mean deviation being that corresponding to the D lines. There are great differences

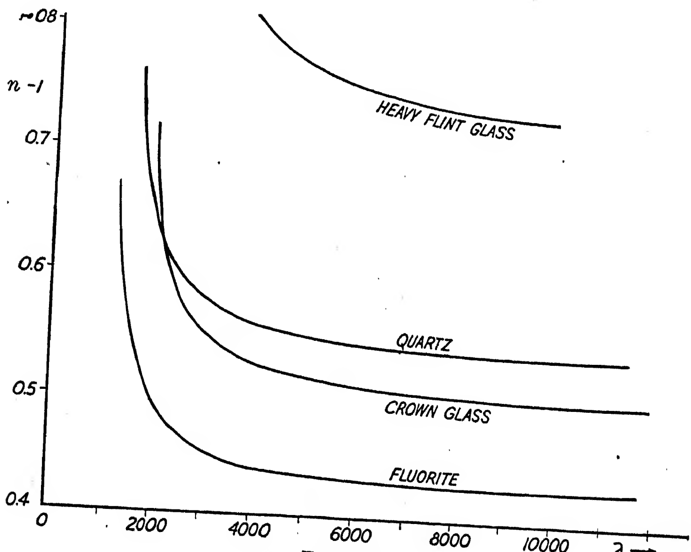


FIG. 540.

between the refractive and dispersive powers of different specimens of glass.

651. Normal Dispersion.—The dispersive power of a given prism (for equal increments of wave-length) varies in different parts of the spectrum, usually increasing toward the violet. Furthermore, when the dispersive powers of prisms of different materials are compared, it is found that they do not show the same variation with wave-length. This will be made clear by reference to Fig. 540, in which are shown curves of $n - 1$ against wave-length for four optical materials. The dispersive power at any wave-length is measured by the ratio of the slope of the curve to its ordinate at that wave-length. That the dispersive powers of different materials do not

bear any simple relation is shown by the fact that one curve cannot be converted into another by a mere change in the scale of ordinates. If this could be done, the spectra formed by different prisms would be alike in the distribution of colors, and one spectrum would simply be a larger or smaller copy of another. Prismatic spectra are not related in this way, and hence such spectra are said to be *irrational.* It is possible, for example, to make a prism of crown glass and one of flint glass which will give spectra of equal length between the lines A and K; but then the positions of the other Fraunhofer lines do not coincide in the two spectra, as they would if the dispersion were rational.

The following table showing the differences between the refractive indices of various substances for the A, D, F, and G Fraunhofer lines illustrates irrationality of dispersion. It will be seen that the ratio $(n_F - n_D)/(n_D - n_A)$, for example, is not the same for the different substances.

	$n_D - n_A$	$n_F - n_D$	$n_G - n_F$	$\frac{n_F - n_D}{n_D - n_A}$
Crown glass	0.00485	0.00515	0.00407	1.062
Heavy flint glass	.01097	.01271	.01062	1.158
Water	.00409	.00415	.00344	1.015
CS_2	.01898	.02485	.02446	1.309

The curves of Fig. 540 are typical of substances which are said to possess "normal" dispersion. Within the limits of the visible

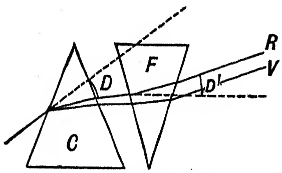


FIG. 541a.

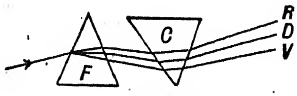


FIG. 541b.

spectrum the relation between index of refraction and wave-length is closely expressed by the empirical relation (Cauchy's formula)

$$n = A + \frac{B}{\lambda^2} + \frac{C}{\lambda^4}$$

where A, B and C are constants varying with the substance. Although a transparent substance may show normal dispersion in the visible region, when the entire spectral range is investigated, the relations are found to be more complicated.

652. Achromatic and Direct-vision Prisms.—The unequal dispersive power of different substances is utilized for making prismatic combinations for producing deviation with very little dispersion (Fig. 541*a*), or dispersion without deviation of the spectrum as a whole (Fig. 541*b*). These two types are respectively called *achromatic* and *direct-vision* prisms.

LENSES

653. Lenses are transparent bodies, generally with spherical surfaces, which form images by changing the curvature of light waves. The ordinary types of single lenses are shown in Fig. 542. The first three forms, known as double-convex, plano-convex, and concavo-convex, are *thicker at the center than at the edges.* If surrounded by a less refractive medium, the central portion of the incident wave is more retarded than the edges by these lenses, and the curvature of the wave is diminished or reversed in direction. These lenses have, therefore a convergent effect. They are called *convex* or *converging* lenses. In the second group, embracing the double-concave, convexo-concave and plano-concave lenses, *the edges are thicker than the center,* so that the outer portions of an incident wave are more retarded than the center. The curvature of the wave is increased and the lenses have a divergent effect. Such lenses are called *concave* or *diverging* lenses. If the two types of lenses are placed in a more refractive medium there is a reversal of these effects.

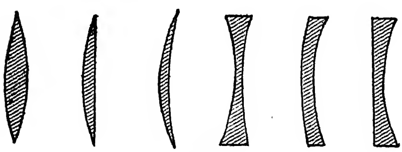


Fig. 542.

654. Equivalent Air Path or Reduced Optical Path.—At a given instant a wave front is in a given position; later it will be in a different position, and may have its orientation and curvature greatly modified by reflection or refraction. The one condition that must always be fulfilled, if the wave is to preserve its identity, is that the time required for the disturbance to travel from a point in the original wave front to the corresponding point in the new wave

front is the same for all parts of the wave. For example, the disturbance traveling from S by the path $SPRQS_1$ (Fig. 543) reaches S_1 at the same time as the disturbance leaving S at the same instant and traveling along $SACES_1$. The latter has been sufficiently retarded by passing through a greater thickness of glass to compensate for the greater distance in air $SPRQS_1$. Similarly, the time required for the wave to travel from P to Q is the same as that from B to D.

In comparing the distances traversed in equal times in different media, account must be taken of the velocity of light in the respective media. For example, in Fig. 543, $PR + RQ = V_1t$; $BD = V_2t$. Therefore, $PR + RQ = (V_1/V_2)BD = nBD$. If BD is the

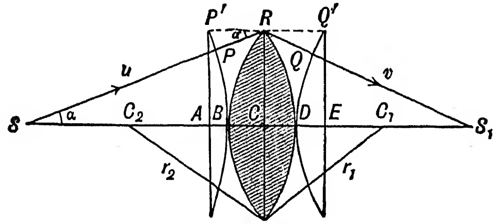


FIG. 543.

distance actually traversed in a medium of refractive index n, the *equivalent air path or reduced optical path* is nBD.

655. Conjugate Focal Relations.—Consider the case of a double-convex lens of refractive index n surrounded by air, the refractive index of which may be taken as unity. Let the radius of curvature of the first surface of the lens be r_1, that of the second r_2. Let u be the distance of the source from the lens. PB is a section of the incident wave front of radius u, and QD that of the emergent wave front, of radius v.

The disturbance actually travels radially from P to R, thence to Q, but if α is very small the path in air may be assumed to be equal to $P'Q'$ without appreciable error. Placing the optical path through the center of the lens equal to this distance, we have

$$P'Q' = AB + BC + CD + DE = n(BC + CD),$$

or

$$AB + DE = (n - 1)\ (BC + CD)$$

Substituting reciprocal radii of curvature for sagittae (§634), this becomes

$$\frac{1}{u} + \frac{1}{v} = (n - 1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right) = \frac{1}{f}$$

In this equation u, v, r_1, and r_2 are considered as mere lengths, that is, numerical quantities without sign, but we shall treat them later as algebraical quantities with signs.

The refraction at the first surface makes the wave less divergent, that is, it tends to converge it toward the opposite side. The same is true of the second surface. Hence both surfaces may be described as converging surfaces. If the curvature of either surface were opposite to its direction in the double-convex lens, it would be a diverging surface.

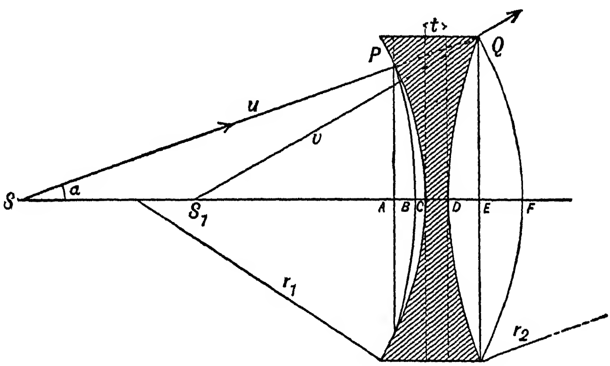


FIG. 544.

If the source of light is at an infinite distance, that is, if the incident waves are plane, $u = \infty$ and $v = f$. Hence f is the distance of the point called *the principal focus*, to which the lens converges plane waves.

In the case of a double concave lens, of thickness CD along the axis (Fig. 544), if the incident wave front is PB and the emergent wave front QF, put the optical path BF equal to the optical path PQ (assumed parallel to the axis, since α is small)

$$AC - AB + n(CD) + DE + EF = n(AC + CD + DE)$$
$$\therefore AB - EF = (n - 1)(-AC - DE)$$

Substituting radii of curvature for sagittae,

$$\frac{1}{u} - \frac{1}{v} = (n - 1)\left(-\frac{1}{r_1} - \frac{1}{r_2}\right) = -\frac{1}{f}$$

Now this equation would be identical with that for the double convex lens if the signs of v, r_1, and r_2 were reversed. Hence, if we consider as positive all quantities (u, v, r_1 r_2 and f) as they appear in Fig. 543, and as negative when they have the opposite direction, the two equations will become identical.

From Fig. 544 it is evident that incident light is made more divergent by the lens. In fact both surfaces are diverging surfaces. The significance of the negative sign of f in the above formula is that the principal focus is virtual, its distance from the lens being measured in a direction opposite to that in which the light actually travels.

By applying the same method to the other types of spherical lenses it will be found that the general solution of all cases is the formula

$$\frac{1}{u} + \frac{1}{v} = (n - 1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right) = \frac{1}{f}$$

provided we adopt the following rule regarding signs:

Consider each of the quantities, u, v, r_1, r_2, f, as positive when it is on the same side of the lens as in the typical case of a double convex lens forming a real image of a real object—negative when on the opposite side.

Negative values of v and f indicate that the light diverges from a virtual focus after passing through the lens. These conventions are consistent with those of §635.

When $u = \infty$, $v = f$, the principal focal distance.

When $u > f$, v is positive and there is a real conjugate focus.

When $u = f$, $v = \infty$. The transmitted beam is parallel.

When $u < f$, v is negative and greater than u for all positive values of u, and there is a virtual conjugate focus.

656. Axes of Lens.—The line passing through the centers of curvature of the surfaces of a lens is called the *principal axis*. In every lens there is a point on the principal axis, called the *optical center*, which has the property that no ray passing through it is deviated in direction, although there is more or less displacement, depending on the thickness of the lens.

The existence of this point may be shown thus: Let two parallel radii of curvature r_1 and r_2 (Fig. 545) be drawn to the two surfaces of a lens. Since the two plane elements of the lens A and A' are parallel, being respectively perpendicular to two parallel lines, the

refracted ray AA' is propagated in a medium with parallel sides and emerges parallel to its original direction. Since the triangles ACC_1 and $A'CC_2$ are similar,

$$\frac{r_1}{CC_1} = \frac{r_2}{CC_2}$$

This is true whatever may be the value of the angle α, therefore C is a fixed point, the optical center of the lens. All ray paths which pass through this point are called *secondary axes.*

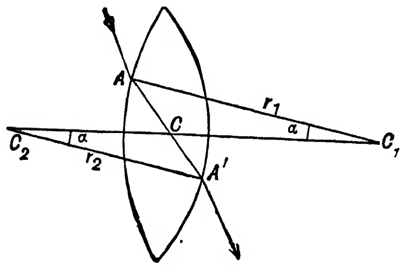


FIG. 545

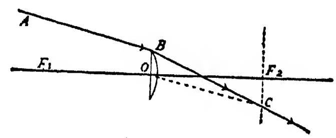


FIG. 546.

In the case of a thin lens, the geometrical center of the lens and the optical center may usually be regarded as coincident.

Path of Oblique Pencil.—If AB (Fig. 546) is a pencil incident obliquely and OC a line drawn from the center of the lens parallel to AB, and intersecting at C the perpendicular plane through F_2 (the *focal plane*), BC is the path of the transmitted pencil, since parallel pencils intersect in the focal plane.

657. Images by Lens.—The image of A (Fig. 547, a, b, c) must lie on the secondary axis AA', that of B on the secondary axis BB'.

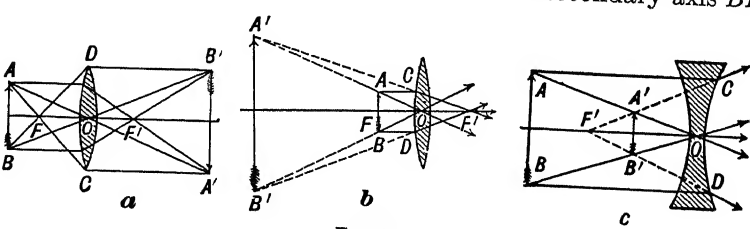


FIG. 547

Rays drawn parallel to the principal axis from the points A and B pass through the principal focus F' and intersect the lines AA' and BB' at the points A' and B', which determine the position and magnitude of the image. Since the point A' lies above the principal axis when the image is on the same side of the lens as the object, and below it when the image is on the other side of the lens, it is

evident that all virtual images formed by a single lens are erect, all real images inverted.

Since an object AB and its image $A'B'$ subtend equal angles from the center of the lens (the angle included between the secondary axes AA' and BB') it is evident that their relative sizes are proportional to their respective distances from the lens, or

$$\frac{AB}{A'B'} = \frac{u}{v}$$

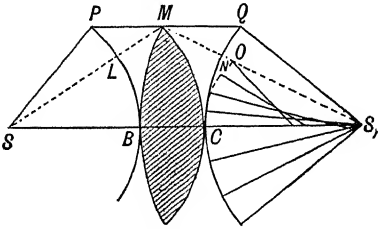


FIG. 548.

658. Spherical Aberration.—In deriving the formula for the conjugate focal relations of lenses it has been assumed tacitly that the emergent wave is spherical. With lenses of small aperture this is shown by experience to be practically true; but when the aperture becomes large there is noticeable spherical aberration. This is illustrated by Fig. 548.

QC represents a spherical emergent wave with every ray converging at S_1 as shown in the lower half of the figure. The actual emergent wave has, however, the shape CO, and the different rays meet the principal axis at different points, forming a caustic curve, as shown in the upper half of the figure.

659. Correction of Spherical Aberration.—If the rays passing through the edge of a lens are stopped by a diaphragm which permits only the central portion of the incident pencil to pass, the spherical aberration will be greatly reduced. It is also possible to grind

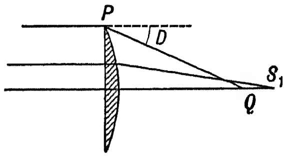


FIG. 549a.

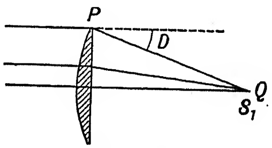


FIG. 549b.

surfaces slightly differing from a spherical form, so that for a *given pair* of conjugate focal distances the emergent wave is truly spherical. Such lenses are called *aplanatic*. In some cases when the conjugate focal distances differ greatly, spherical aberration may be reduced by making the two surfaces of the lens of different curvatures. Consider, for example, a plano-convex lens of great aperture first with the plane face (Fig. 549a) then with the convex face (Fig.

549b) toward a source so distant that the incident light is parallel or nearly so. If we consider the deviation of the ray PQ in each case, it is evident, on recalling the condition of minimum deviation by a prism, that in the second case the angle D will be less than in the first, because the refracting edge of the lens is then more nearly in the position with respect to the incident and emergent rays which gives minimum deviation, and consequently the nearest approach of the ray PQ to the focus S_1. The above is a special case of the more general rule that when the deviation produced by two surfaces is quite different, spherical aberration will be greatly reduced by redistributing the deviation equally between the two surfaces.

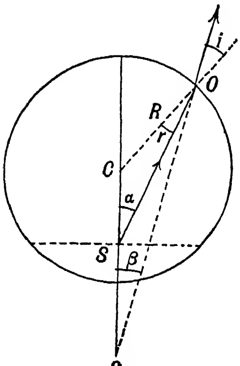


FIG 550.

One form of thick lens of great angular aperture commonly used as part of microscope objectives is almost entirely free from spherical aberration. This has the form of a hemisphere, and is placed with its plane surface toward the object (Fig. 550). In the method of "oil immersion" the space between the lens and the object is filled with an oil of the same refractive index as the lens. If R is the radius of the hemisphere and n the refractive index, an object placed at a distance R/n from the center of the spherical surface will produce a virtual image at a distance nR free from all spherical aberration.

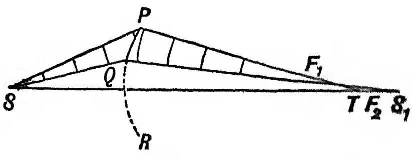


FIG. 551.

660. Focal Lines.—If a pencil of light falls obliquely on a converging lens, two real focal lines will be formed, like those due to a concave mirror. If the lens is divergent, these focal lines will be virtual. The formation of these lines by a converging surface is made clear by considering the effect of a single refracting surface PQ, imagining it to be extended to R, so that SS_1 is the principal axis (Fig. 551) The rays transmitted through a small surface of which PQ is a section will pass through a narrow arc formed by revolving F_1 a short distance about SS_1. This is the *primary focal line.* These rays will all intersect the axis SS_1 between S_1 and T. The normal cross-section of this pencil is a narrow lemniscate-shaped region at F_2, the *secondary focal line,* at right angles to the primary focal line. The second refracting surface of the lens will modify but not change the general character of this result.

661. Cylindrical Lens.—The effect of such a lens is like that of a cylindrical mirror. A point source S has two linear images, as shown in Fig. 552, one AB parallel and the other $A'B'$ at right angles to the axis of the lens. The image that is parallel to the axis may be either real or virtual; the other, $A'B'$, is virtual and may be considered as due to each longitudinal strip of the lens acting as a prism of the same angle. Any lens with different curvatures in planes at right angles to each other will give similar focal lines or astigmatic images.

662. Combination of Lenses.—Consider a thin lens of focal length f_1, with an object distance u, forming a real image at a distance w. Then

$$\frac{1}{u} + \frac{1}{w} = \frac{1}{f_1}$$

If a second thin lens of focal length f_2 is placed so close to the first that the distance between their optical centers is negligible, *converging* light from the first lens strikes the second lens. Hence the second lens has a *virtual object* at a *negative* distance w equal to the *positive* image distance of the first lens. Therefore if v is the image distance for the second lens

$$\frac{1}{-w} + \frac{1}{v} = \frac{1}{f_2}$$

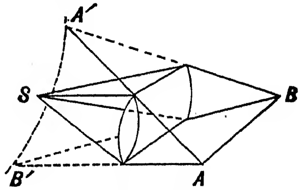


FIG. 552.

From these two equations we get

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f_1} + \frac{1}{f_2} = \frac{1}{f}$$

Hence the two lenses are equivalent to a single lens of focal length f. The relation between f_1, f_2 and f is general, provided the proper signs are given to f_1 and f_2.

The same method of procedure may be applied to find the position of the final image in the case of two lenses separated by a distance d. In this case the object of the second lens is at a distance $-(w - d)$.

663. Chromatic Aberration.—Since

$$\frac{1}{f} = (n - 1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right) = (n - 1)K,$$

it is evident that the principal focal distances are different for different colors, being less for violet than for red (Fig. 553). There is no way to remedy this defect in a single lens, but it may be greatly reduced by a suitable combination of lenses.

By combining two or more lenses of different dispersive powers two or more given colors may be brought to the same focus (Fig. 554), just as prisms may be combined to give deviation without dispersion (Fig. 540). If two thin lenses in contact are used,

$$\frac{1}{f_1} + \frac{1}{f_2} = \frac{1}{f}$$

for each color. If we wish to combine the two colors corresponding to the C and F lines, f must be the same for both.

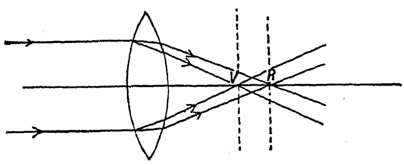


FIG. 553.

Chromatic aberration may also be reduced by using two lenses of the same index of refraction at a certain distance d from each other. To take a specific case, if the second lens is placed at a distance from the first equal to its own focal length, the rays of different colors which diverge from each other at the first lens will be made approximately parallel by the second. If an object is placed at the principal focal point F (Fig. 555) of the combination, a virtual image at infinity will be formed, and, as shown by the figure, the violet and the red images will subtend approximately equal angles α at the eye, and will, therefore, be superimposed on the retina.

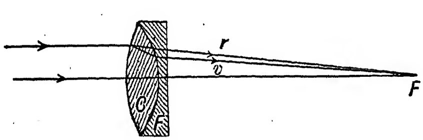


FIG. 554.

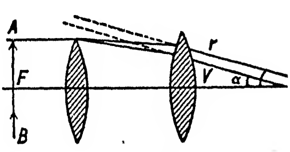


FIG. 555.

664. Calculations of Achromatic Combinations.—If n_F' and n_C' be the refractive indices of the first lens, n_F'' and n_C'' those of the second,

$$\frac{1}{f} = \frac{1}{f_F'} + \frac{1}{f_F''} = (n_F' - 1)K' + (n_F'' - 1)K''$$
$$= \frac{1}{f_C'} + \frac{1}{f_C''} = (n_C' - 1)K' + (n_C'' - 1)K''$$

therefore,

$$(n_F' - n_C')K' = (n_C'' - n_F'')K''$$

The values of $K' = 1/r_1' + 1/r_2'$ and $K'' = 1/r_1'' + 1/r_2''$ may be arbitrarily chosen to satisfy this relation. Since $n_F > n_C$ it is evident that K' and K'' must be of opposite sign, so that either f' or f'' must be negative. If f'' is negative and greater than f', f is positive and the lens is convergent. If f''

negative and less than f' the combination is divergent. Usually the positive ns is of crown glass, the negative of flint, and they are shaped to fit close ogether, so that $r_2' = -r_1''$ and often $r_2'' = \infty$ (Fig. 554).

ADDITIONAL REFRACTION PHENOMENA

665. Total Reflection.—If a ray of light travels from a more to less refractive medium, the angle of emergence i is greater than he angle of incidence (which, being in the more refractive medium, ay still be called r for consistency). Since $\sin r = (\sin i)/n$, nd since i has a maximum limit of 90 degrees, r has a maximum mit k such that $\sin k = 1/n$. No pencil incident on the boundary

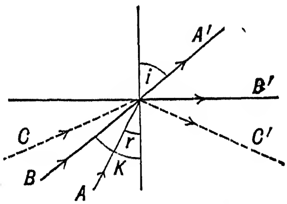


FIG. 556.

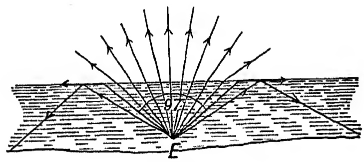


FIG. 557.

t a greater angle than k, the *critical angle*, can emerge. It will, herefore, be totally reflected (CC', Fig. 556). Since $\sin k$ varies nversely as the index of refraction the critical angle is different or different colors. Violet will first be subject to total reflection s r increases, and finally the red.

A parallel-sided plate cannot be used to show total reflection, ince any pencil entering such a plate must emerge at the same ngle. Objects of prismatic form are best adapted for the purpose. The effect may be seen by looking down through the side of a glass ontaining water or at a test-tube sunk in water. A fish can see bjects throughout the space above the water, but he sees them listorted into the limited cone of angle $2k = 97°$ (Fig. 557).

Some values of k are given below:

Water	48° 36′	Quartz	40° 22′
Crown glass	43° 2′	Diamond	24° 26′
Flint glass	37° 34′		

The smaller the critical angle of a jewel with regular facets, the greater the proportion of light totally reflected by it. This explains the great brilliancy of the diamond.

The index of refraction of a liquid or of a small portion of an opaque object may be determined by measuring the angle of total

reflection from its surface when in contact with a more refractive medium and using the relation $\sin k = n/n_1$, where n is the index of the less and n_1 that of the more refractive medium.

666. Transition Layer.—It seems quite possible that the change of index of refraction at the boundary is not abrupt, but that there is a transition layer t due to interpenetration of the two media, or occlusion at the surface causing a gradual change in the index. If this be the case, total reflection may be considered as altogether due to refraction When the angle of incidence is equal to or greater than k, the wave front in the transition layer will swing around and become normal to the surface (Fig. 558): then the lower edge will gain on the upper and the wave will swing back into the first medium If we consider an air film between two refracting media the two transition layers may encroach on each other (Fig. 559), in which case the lower edge of the

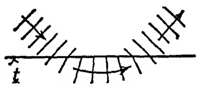


FIG 558

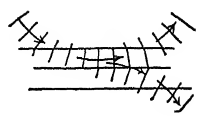


FIG 559

wave will be retarded, and a part of it will pass into the third medium. It might be expected, therefore, that if the air film from which total reflection takes place is very thin total reflection will cease. This has been found to be the case If a right-angled total reflecting prism with a slightly convex hypothenuse surface is pressed against a glass plate, total reflection takes place from the hypothenuse of the prism when the angle of incidence i is sufficiently large, but some light will always be transmitted through the region surrounding the point of contact even where the air film has a measurable thickness. It is found that the thickness of the air film through which transmission can occur (which may be considered as approximately the thickness of the transition layer) differs with the wave-length and with the angle of incidence, and may reach several thousandths of a millimeter

667. Mirage.—Examples of the type of total reflection referred to above are found in the case of refraction by gases of varying density. This phenomenon is called mirage. The air above a furnace or a heated surface such as a pavement exposed to the sun's rays rapidly cools in going upward and therefore increases in density and refractive power. If the line of vision forms a small angle with the surface, distant objects are seen apparently reflected from the surface. The formation of one type of mirage is shown in Fig. 560. The object AB is viewed directly through the pencils OA, OB, while an inverted image $A'B'$ is also seen, due to the refraction of the pencils OEA, OFB, by the heated air near the ground. This is one of several types of atmospheric mirage. Other types

showing distortion or displacement of objects are due to local differences of temperature in the atmosphere, which cause changes in density and refractive power. They are very easily seen by viewing objects at a grazing angle across heated surfaces. Similar effects are to be seen by looking through sheets of glass with irregular surfaces, or non-homogeneous mixtures of liquids, such as water

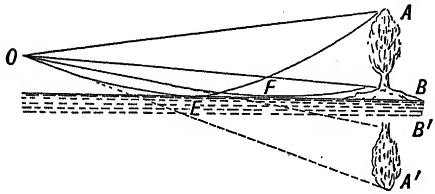


FIG. 560.

with an excess of salt crystals at the bottom of the vessel, or with alcohol above and imperfectly mixed with it.

When the sun is near the horizon, the rays reaching the eye traverse strata of air of gradually increasing density, which cause them to bend downward. For this reason the sun is visible when it is actually below the horizon a distance about equal to its own diameter.

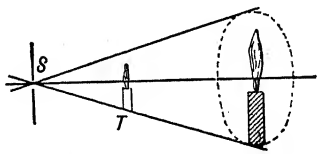


FIG. 561.

The twinkling of stars is due to a similar cause. Their apparent direction and intensity are subject to rapid fluctuations as masses of air of varying density drift across the line of sight; both brightness and color vary, as can be seen by shaking the head so as to elongate the image.

If sunlight be focused on a small hole beyond which it diverges to a screen, a jet of coal gas, hydrogen, or carbon dioxide will cast a clear image on the screen. The difference between the refractive power of the jet and that of the air will cause an alteration in the distribution of light on the screen, which will make the projected areas lighter or darker than the surrounding space. Ether vapor poured from a beaker, or ether vapor vortex rings, may thus be made visible. If a Bunsen burner be placed in the cone of light a beautiful representation of the flame and the currents of heated gases will be formed on the screen. The photography of sound waves uses the same principle (§588).

668. The Rainbow is a bright arc showing the spectral colors, due to the sunlight refracted by raindrops. Sometimes several bows are seen, the inner or primary bow being always the brightest,

and all being arcs of circles with centers on the prolongation of the line passing from the sun through the observer. The primary bow is violet on the inside, red on the outside; in the secondary bow the order of colors is reversed.

If parallel rays are incident on the upper half of a refracting sphere they will be in part refracted, internally reflected, and transmitted downward as shown in Fig. 562. Rays will also enter the lower half, and there will be multiple reflection within the sphere, but for the present we shall fix our attention upon the rays which reach the eye at O after one internal reflection. As indicated by

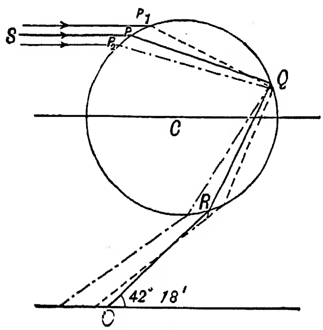


Fig. 562.

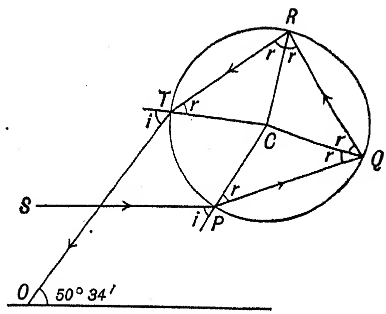


Fig. 563.

the course of the rays incident at P, P_1, and P_2, there is an angle of minimum deviation, below which no rays once internally reflected pass. All the rays emerging at nearly this angle are parallel or nearly so, and therefore their intensity varies little with the distance from the drop, while rays emerging in other directions are widely divergent.

As n varies with the color, the minimum deviations D are different for the various colors. In the primary bow the minimum deviation of the red is 137° 42′; of the violet, 139° 37′. In the secondary bow the corresponding angles are 230° 34′ and 233° 56′.

From Fig. 562 it appears that light will be received by the observer at O from all the raindrops lying in an arc subtending an angle 180°—D with the axis passing from the sun through the observer's eye. In the primary bow this angle is 42° 18′ for the red and 40° 23′ for the violet, so that the bow will be bordered with violet on the lower side, red on the upper. The secondary bow is due to rays incident on the lower half of the drop, twice internally

reflected, and then transmitted downward, thus inverting the order of the colors (Fig. 563). The angle subtended by this bow is $D - 180°$, or 50° 34′ for the red, 53° 56′ for the violet.

An artificial rainbow may be made by causing a beam of sunlight to fall on a spherical vessel filled with water, through an opening in a screen. The interior of the circle reflected on the screen is illuminated by the scattered light which has been once reflected, while the space between the primary and secondary bow is quite dark.

INTERFERENCE

669. Examples of Interference.—Fresnel, a young French artillery officer, about 1815, produced effects similar to those described

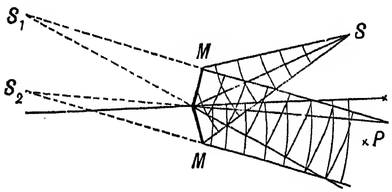


FIG. 564.

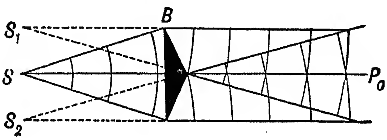


FIG. 565.

in §625 by light diverging from a slit S and reflected from two adjacent mirrors MM inclined at nearly 180° so as to be almost in the same plane. As shown by Fig. 564, the light arriving at any point P where the pencils overlap appears to come from the two virtual sources S_1 and S_2, the effect of which is precisely the

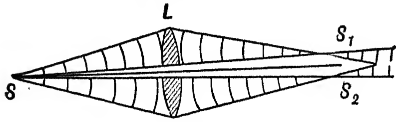


FIG. 566.

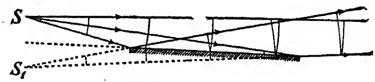


FIG. 567.

same as that of the two real sources in Young's experiment. (The term virtual source applies to a point from which the waves appear to diverge without really originating at that point.)

Fresnel also produced interference effects by the use of a biprism B equivalent to two prisms of very small refracting angle placed base to base (Fig. 565). Here, again, it is evident that the transmitted light appears to come from the two virtual sources S_1 and S_2.

Another method of obtaining similar interference effects is by means of a convex lens L cut along a diameter in two halves which are slightly separated giving two real or virtual images of the source, from which the waves diverge and overlap. This is known as the Billet split lens (Fig. 566).

The interference effects due to Lloyd's single mirror (Fig. 567), are caused by waves coming respectively from the real source S and the virtual source S'. The fringes are easily obtained by reflecting the light from a narrow slit or lamp filament at grazing incidence from a mirror of black glass, in order that the effects may not be complicated by reflection from the rear surface.

670. Newton's Rings. Robert Boyle described the brilliant colors observed in soap bubbles and other thin films, an effect which appeared to depend solely on the thickness of the films, not on their nature. Newton investigated this phenomenon, which he tried, with poor success, to explain in terms of the emission theory. In order to secure a thin film of air, varying in thickness in a determinate manner from point to point, he pressed a convex glass lens of great radius of curvature against a piece of plane glass. If light falls normally on such a combination, light of a given color is found to be reflected in a greater proportion than the other colors from all points where the film has a given thickness, the predominant color varying with the thickness. As the loci of points of equal thickness form circles about the region of contact, colored rings are observed concentric with this point. These have been called Newton's rings, or the *colors of thin plates*. Colored rings are likewise observed in the transmitted light. These are not so brilliant, however, as those due to the reflected light, as the transmitted colors are mixed with a large proportion of unmodified white light. The colors in the two sets of rings are complementary—that is to say, the light transmitted through a given point is white deprived of the color which is most strongly reflected from that point. If monochromatic light is used the rings are alternately dark and of the color used. In a wedge-shaped film these bands are parallel to the edge of the wedge; in a film of uniform thickness circular bands are produced under certain conditions, uniform color effects under others. These colors of thin plates are seen in all kinds of thin transparent films, such as soap bubbles, films of oil on water and thin sheets of mica.

671. Explanation of Newton's Rings.—Thomas Young showed that the colors of thin films can be explained very simply as a result of the interference of waves reflected from the two surfaces of the film. Let a ray from a point in an extended source be incident on the upper surface of the film at A (Fig. 568), and for simplicity imagine the two surfaces to be plane and parallel. A small part will be reflected at A, the remainder being transmitted to B where it is again subject to reflection and refraction by the lower surface. The process is repeated at C, D, E, etc., but the components become very weak after a few

reflections. The reflected or transmitted light may be brought to a focus at S_1 or S_2 by a lens (which may be the eye), and a maximum or a minimum of intensity may be produced at these points, according to the phase differences between the component rays.

We shall now find the difference in path for the first two rays leaving the upper side of the film. Let n be the index of refraction of the film and n_1 that of the surrounding medium. To reach the wave front CP, the light reflected from A travels the distance AP in the first medium, while the interfering component has to travel the distance $AB + BC$ in the film. The wave would travel a distance n_1AP in air while traveling the distance AP in the medium of index n_1, and the distance $n(AB + BC)$ in air while travelling the distance $AB + BC$ in the film (§654). Hence the equivalent difference of path in air, or the optical difference of path is

$$d = n(AB + BC) - n_1AP$$

But $AC \sin i = AP$ and $AC \sin r = CQ$; therefore

$$AP = CQ\frac{\sin i}{\sin r} = \frac{n}{n_1}CQ.$$

Hence

$$\begin{aligned} d &= n(AB + BC - CQ) = n(AB + BQ) \\ &= nAR \cos r = 2nt \cos r. \end{aligned}$$

This is also the path difference for succeeding pairs of rays.

In Newton's rings the film is of air, so we may put $n = 1$ and $d = 2t \cos i$, where i is the angle of refraction in air. The principal change introduced by the fact that the surfaces are not exactly parallel is to cause the reflected rays to diverge slightly. The eye or observing telescope must then be focussed on a point near the film itself, instead of on infinity, to obtain the most distinct effect.

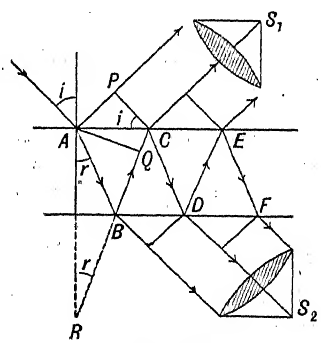


FIG. 568.

672. Phase Changes Due to Reflection.—It is evident that, so far as geometrical differences of path are concerned, there should be reenforcement from the components reflected from the region of contact, where the thickness of the film is so small compared with the wavelength of light that it may be ignored. As a matter of fact, the center of the reflected system of fringes is black. Young inferred by analogy that at the boundaries of different media light waves are subject to changes of phase similar to those observed in the case of material waves (§252) so that waves incident from air on a more refracting medium may behave like waves of sound reflected from a medium denser than air, while a light wave traveling in the opposite direction will behave like sound waves emerging from the free end of an organ pipe. The waves reflected from the upper surface of the air film pass from a more to a less refractive medium; at the lower surface the contrary is the case. If t is small compared

with the wave-length, there should be a difference of half a period introduced in the act of reflection, which will cause destructive interference. The transmitted components have a difference of phase of an entire period caused by two internal reflections, and therefore will be concordant. This would explain the black spot seen in the center of the reflected system of Newton's rings. It is also observed that soap films as they get thinner run through a brilliant series of colors, when viewed by reflected light, finally becoming black just before they break.

If the lens is of crown glass, the plate of flint glass, and if the interspace is filled with a liquid of intermediate index of refraction, such as oil of cloves, the central spot of the reflected system will be bright, that of the transmitted system dark. This confirms Young's theory.

673. Maxima and Minima in Newton's Rings.—When Newton's rings are produced by an air film, the condition for a maximum of given wave-length λ in the reflected light is (remembering that a loss of half a period in reflection is equivalent to a path difference of $\lambda/2$),

$$N\lambda + \tfrac{1}{2}\lambda = \frac{2N+1}{2}\lambda = 2t \cos i$$

and for a minimum,

$$N\lambda = 2t \cos i$$

In the transmitted light the maxima are given by

$$N\lambda = 2t \cos i$$

and the minima by

$$\frac{2N+1}{2}\lambda = 2t \cos i$$

In the above expressions N is the ordinal number of the rings counted from the center.

674. Plate with Parallel Surfaces.—If the image of an extended source of monochromatic light such as a sodium flame reflected in a very thin plate be examined with the eye, it will be found to be crossed by dark and bright interference bands. The bands are curved, their loci being given by $\cos i$ = constant, since each corresponds to a particular value of the path difference $2d \cos i$. Such bands are known as Haidinger's fringes, or fringes of equal inclination. With thicker plates the fringes will not be seen unless the light is very nearly monochromatic, and unless they are observed in a direction almost perpendicular to the plate. At other angles, Fig. 568 shows that with a thick plate not more than one ray can enter the eye at a time, and hence interference is impossible.

675. Stationary Light Waves.—Stationary waves (§253) may be expected if plane waves of light are reflected normally from a mirror, but as the distance between the nodal planes is only $\lambda/2$ and light waves are very short, it is difficult to verify their existence. Wiener did so by a very ingenious device. A glass plate AB (Fig. 569) was covered with a very thin photographically sensitive collodion film, and placed film downward over a silvered mirror MN with a very slight inclination between the two surfaces. After exposing the

plate to a beam of light incident normally on the mirror, it was developed, and dark bands were found in the film, running parallel to the line of intersection of the two surfaces, as indicated by the shading below MN. From the figure it is clear that the sensitive surface crossed the nodal and anti-nodal planes in such a manner as to produce this effect. At the points in contact with the mirror no effect was produced in the film. This proved the existence of a nodal plane at that surface, as in the case of the analogous sound experiment. This is the basis of a system of color photography invented by Lippmann.

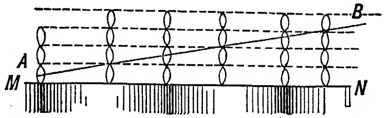


FIG. 569.

DIFFRACTION

676. Observation of Diffraction.—If light from a small source or aperture passes by the edge of an obstacle and falls on a screen, it is found that the illumination gradually fades away in the geometrical shadow, while outside the shadow a series of colored bands

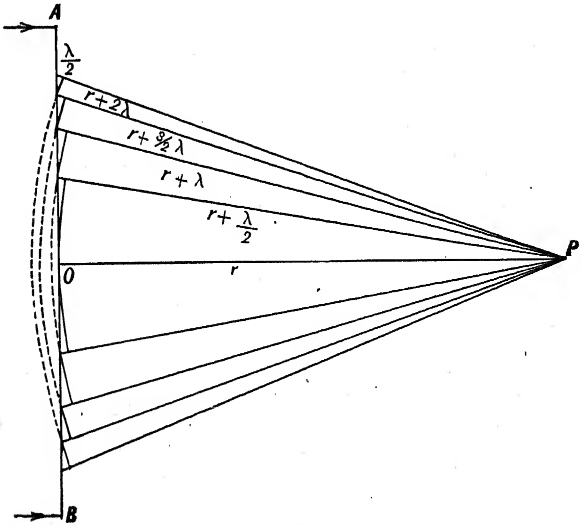


FIG. 570.

appears. If a card or knife blade is held between the eye and a distant source of light it will be found that the red light is most deflected into the shadow, the violet the least, so that a short spectrum is formed. Such phenomena are examples of what is known as **Diffraction.** They are, in fact, interference phenomena

between wavelets coming from adjacent points of the same wave front

677. Diffraction by a Circular Opening or Obstacle.—Let us find the effect of an extended plane wave front AB (Fig. 570) at the point P. In accordance with Huyghens' principle, the resultant effect at P may be regarded as the sum of the effects separately due to all the points in the wave front, each originating its independent set of wavelets. Waves of different lengths must be separately considered in this analysis. If $OP = r$, describe about P as a center spheres of radii $r + \lambda/2$, $r + \lambda$, $r + 3\lambda/2$, etc. These spheres will in intersect the wave front circles, as shown in Fig. 571, concentric with O, the *pole* of the wave with respect to P. The areas between successive circles are called *half-period zones*.

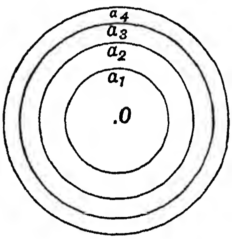


FIG. 571.

The student may easily find by calculation that these areas are approximately equal. It is evident that the disturbances originating in all points on a circle about O will reach P at the same time, and that the average phases of the resultant effects at P of successive zones will differ by half a period. The amplitude produced by each at P slowly diminishes as its radius increases, on account of increasing obliquity. The distance of each zone from P also increases, but the effect due to this is practically counterbalanced by a slight increase in the area of successive zones. The total amplitude produced by the wave at P is, therefore, the algebraic sum of a series of terms slowly diminishing in magnitude and alternating in sign (direction of displacement). If a_1, a_2, a_3, etc., are the amplitudes at P due to the central area and successive zones, and A the resultant amplitude,

$$\begin{aligned} A &= a_1 - a_2 + a_3 - a_4 + a_5 \cdot \quad \cdot \pm a_n \\ &= \tfrac{1}{2}a_1 + \tfrac{1}{2}(a_1 - 2a_2 + a_3) + \tfrac{1}{2}(a_3 - 2a_4 + a_5) \cdot \cdot \cdot \pm \tfrac{1}{2}a_n \end{aligned}$$

As the successive values of a differ very slightly from each other and diminish in accordance with a regular law, the quantities in parentheses and a_n are very nearly equal to zero. Therefore, to a close approximation,

$$A = \tfrac{1}{2}a_1$$

or the amplitude at P due to the whole wave is one-half and the intensity one-fourth that due to the central element if it alone

were effective. If the whole wave except the central element is covered, the illumination at P will actually be increased, the amplitude in that case being a_1. If all but the two central elements

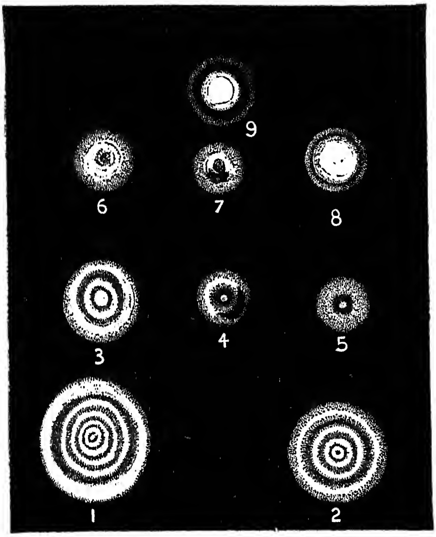


FIG. 572.

are covered the effect at P is $A = a_1 - a_2 = 0$ nearly. If three elements are uncovered, $A = a_1 - a_2 + a_3 = a_1$ nearly. These conclusions are easily verified by experiment. If small circular openings of different sizes be placed in a pencil of light diverging from a pinhole, maxima and minima will be found in the centers of the bright areas projected on a screen through the openings (Fig. 572). These holes decrease regularly in size from 1 to 9. If the screen be moved (thus changing the number of effective half-period elements subtended by the holes at the screen) maxima change to minima and *vice versa*, or if white light is used the bright spot at the center changes color. The central spot is surrounded by a series of colored bands of similar origin but not so easily explained by elementary methods. If a hole is smaller than the first two half-period elements, there are no maxima and minima within the illuminated area on the screen, as there can be no possible complete

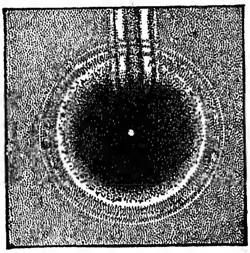


FIG. 573.

cancellation in the wavelets coming through the hole, and consequently a diffuse circular patch of light is cast on the screen, which increases in size as the opening is made smaller.

If a small disk be placed in the path of the light, so as to cover a few half-period elements as viewed from P—say three—the amplitude at P will be $A = a_4 - a_5 + a_6 - a_7 \cdots = \frac{1}{2}a_4$. A bright spot will therefore be seen at the center of the shadow, nearly as intense as though the disk were removed. At adjacent points not on the axis there will be discordance of phase between the disturbances coming around the edge of the disk, resulting in destructive interference.

This experiment may be performed by mounting an accurately circular disk several millimeters in diameter on a piece of glass plate and placing it in the

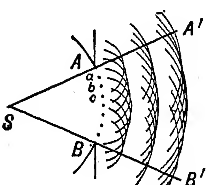


FIG. 574.

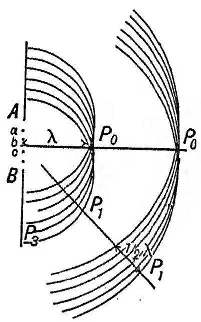


FIG. 575.

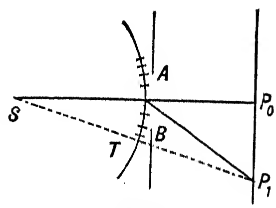


FIG. 576.

pencil of sunlight from a small pinhole opening several meters away. The bright spot in the center of the shadow may then be seen on a screen a few meters beyond the disk. A reproduction of a photograph of this effect is shown in Fig. 573.

678. Waves through Large Opening.—The points a, b, c, etc., in the wave front AB (Fig. 574) act as centers of disturbance and propagate wavelets to the tangent surface $A'B'$. It is evident from the figure that only those wavelets between the lines AA' and BB' interfere in such a way as to produce a regular wave front. Outside of these lines the wavelets cross each other in all directions and in all possible phases at random, so that the resultant disturbance is zero except in the immediate neighborhood of AA' and BB' where diffraction effects are produced.

679. Waves through Small Opening.—From Fig. 575 it is clear that no opposition of phase between the elementary wavelets from a, b, c, etc., can arise until the point P_1 is reached, where the difference between AP_1 and BP_1 is half a wave-length, and even then the disturbances from the extreme points A and B alone are in opposite phase. Only when this extreme difference of path is a whole wave-length can complete destruction arise. In this case

we see that the disturbances from two halves of the opening reach P_1 with an average difference of path of half a wave-length, so that the wavelets nullify each other pair by pair. If the opening is less than a wave-length in width some effect is produced even at the point P_3. The effect is evidently always greatest at P_0, where the wavelets meet very nearly in the same phase, and least at P_3 where there is the greatest diversity of phases.

680. Narrow Slit.—If two straight edges are opposed so as to form a narrow slit of width AB (Fig. 576) there will be a bright band at P_0 if only the first half-period element of each half of an incident wave are exposed. If two on each side are exposed the effect at P_0 is

$$A = 2a_1 - 2a_2 \text{ (nearly zero)}$$

If three half-period elements on each side are exposed

$$A = 2a_1 - 2a_2 + 2a_3 \text{ (maximum)}$$

Thus there will be successive maxima and minima at P_0 as the slit is widened. If the slit subtends two or any even number of half-period elements as viewed from P_1, a point off the axis (T being its pole), they will neutralize each other in pairs; if it subtends an odd number of such elements, there will be destructive interference between pairs, leaving the odd one effective, consequently there will be a series of maxima and minima on each side of the axis.

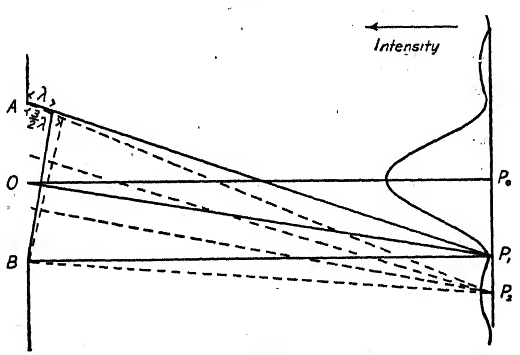


FIG. 577.

681. Width of the Bands Formed by a Narrow Slit.—If $AP_1 - BP_1 = \lambda$ (Fig. 577) we may consider the effects at P_1 of AO and OB to be nearly the same numerically, but to differ in average phase by half a wave-length. The two almost cancel each other. At P_2, where $AP_2 - BP_2 = \frac{3}{2}\lambda$, we may imagine the slit divided into three nearly equal strips, which contribute effects at P alternating in phase. Two cancel each other, leaving the third effective.

The curve gives the variation of intensity on the screen. The pattern shown is obtained only when the slit-width AB is very much smaller than the distance D between the screen and the slit. Under these conditions it can be shown as in §625 that the width of the central maximum

$$2P_0P_1 = \frac{2D\lambda}{AB}$$

The other bands are of half this width, or $D\lambda/AB$. The width of all the bands is, therefore, inversely proportional to the width of the slit. The central

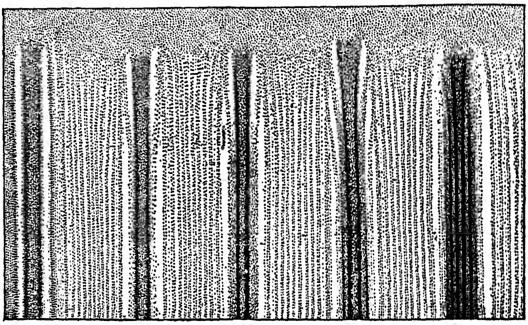


FIG. 578.

maximum is white, but the secondary maxima are colored, since P_0P_2 is different for each different color. These effects may be observed by allowing light from a narrow slit to pass through a second adjustable slit and fall on a screen, or more simply by looking through a narrow slit or the space between two fingers at a distant light.

Within and close outside the shadow of a wire or needle cast by a linear source similar fringes are observed. (See Fig. 578, showing shadows of needles of different sizes.)

682. Resolving Power.—If the light from a narrow slit passes through another slit to a screen the central maximum may be regarded as an image of the

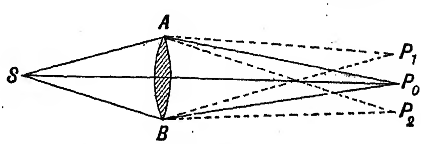


FIG. 579.

first slit (corresponding to a pin-hole image). The wider the second slit is opened (up to the point where diffraction effects cease) the narrower and sharper this image will be. Similar considerations apply to light from point sources through circular openings. If we look through a small pinhole at a distant light it will appear much larger than when viewed with the naked eye. The filament of a lamp appears thicker when seen through a narrow slit. If an image is formed by a lens or mirror the same conditions hold as for a narrow slit, the

lens or mirror preserving the uniformity of phase of the whole wave with respect to the focus that exists for a narrow slit with respect to its central maximum. Consider the image of a narrow source S (Fig. 579). At P_1 and P_2 on each side of P_0 there will be a minimum if $AP_2 - BP_2 = \lambda = BP_1 - AP_1$ in which case the disturbances from the two halves of the lens reach P_1 and P_2 in opposite phases and cancel each other. The width of the image, which is merely a diffraction maximum, is therefore, P_1P_2. If the source is very small, the central maximum will be of the shape of the opening, but differently oriented, because any particular dimension in the image will be inversely proportional to the same dimension in the opening, as appears from the relation (§681)

$$P_1P_2 = 2P_0P_1 = 2D\lambda/AB$$

Observation shows that two diffraction maxima cannot be clearly separated if they are closer than the distance from a maximum to the adjacent minimum. Fig. 580 shows the intensity curves for two diffraction images in this position.

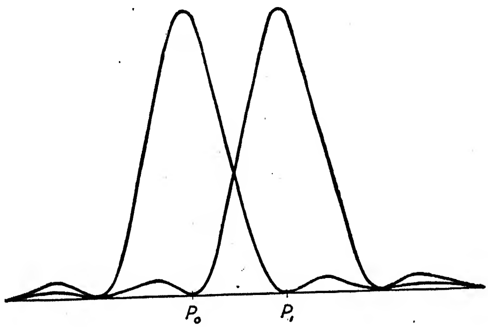


FIG. 580.

The sum of the two intensities at the center is about four-fifths of that at either maximum. If the image of a second source S' lies at P_1 (Fig. 579) it can barely be seen as separate from P_0, the image of S. The two images will thus overlap if the angle subtended by the objects at the lens is less than $\alpha = \lambda/AB = P_0P_1/D$. This is called *the minimum angle of resolution.*

The formulas just given apply strictly only to the case of a long rectangular opening of width AB. For a circular opening of width AB, we must replace λ by 1.22λ.

The preceding applies only to small sources or objects. If the source is large each point will have a diffraction maximum at the focus. These maxima will overlap and blot out diffraction effects except at the boundaries of the image.

Stars are practically point sources of light. Their images when formed by a telescope with a small objective appear much larger than when formed with a large telescope, the diameters of the central maxima being inversely proportional to the diameter AB of the lens. The image of a double star formed by a small telescope may be one large blur, while that formed by a large telescope consists of two distinct points of light. The ability to separate the

images of two small adjacent sources is called resolving power. Since it is evidently inversely proportional to the angle of minimum resolution, it is directly proportional to the diameter of the lens, mirror, or prism forming the image—or to the cross-section of the effective beam of light, if it does not fill the aperture.

If we had larger eyes we could see much finer details than we now do. On the other hand, if we look through a small pin-hole at a distant light it will appear much larger than when viewed with the unaided eye. For a similar reason it is physically impossible for small insects to see details clearly. To them an incandescent lamp filament must appear as it does to us when we look at it through a very small pin-hole.

683. Diffraction Grating.—If there are a number of narrow and equidistant parallel openings in a screen, each pair of openings will produce effects similar to those observed in Young's double

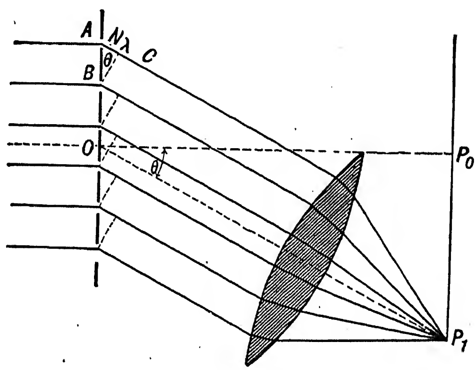


FIG. 581.

slit experiment. If a lens is placed in front of such a diffraction grating, as it is called (Fig. 581), the same path difference will exist between any pair of adjacent parallel rays. If $a = AB$ is the distance between openings and if the angle between OP_0 and OP_1 is θ, the difference of path between corresponding parts of adjacent slits (*e.g.* A and B) is $AC = a \sin \theta$, and the condition that there shall be a maximum at P_1 for the wave length λ is

$$a \sin \theta = N\lambda$$

If light of one wave-length is used, there is a series of maxima on either side of the axis in positions where $a \sin \theta$ equals 1, 2, 3, etc., wave-lengths.

If white light is used, corresponding maxima for two different colors are at different distances from the axis. The central maximum P_0 is white, as the condition for reenforcement at that point

($N = 0$) is the same for all colors. The other maxima are drawn out into spectra on each side of the axis, as θ varies with the wave-length. The value of the ordinal number N determines the order of the spectrum.

If θ is a small angle, the distances between points in the spectra are nearly proportional to the differences of the corresponding wave-lengths, so that the spectra formed by gratings are said to be *normal*, as contrasted to those due to prisms, in which there is no simple law of distribution. All gratings give spectra which are alike in their distribution of colors, although they may differ in length. The lengths of the spectra increase directly as the order of the spectrum, so that those beyond the first overlap, and they also rapidly diminish in intensity.

A grating such as has been considered above is called a *transmission* grating. It consists of lines ruled (several thousand to the inch) on glass by a diamond point or a photographic replica. *Reflection* gratings are made by ruling lines on a polished surface of speculum metal or aluminum; the incident light is reflected by the polished strips between the rulings.

The effect due to a grating, so far as position of maxima is concerned, depends on the separation of the openings but is independent of the number of openings. The intensity, however, increases with the square of the number, the amplitude of vibration being in proportion to the number of openings. Thus the intensity due to a 20,000-line grating is 10^8 times as great as that due to two slits.

The resolving power of a grating is also greater. The width of a maximum in the interference bands given by two slits is (§625) $w = D\lambda/a$. The width of the maxima given by a grating having n openings is $w = D\lambda/na$, since na is the aperture of the grating, so that this width w is inversely as the breadth of the grating. Diffraction gratings are generally used for measurements of wave-length.

A grating with crossed lines gives a beautiful series of crossed spectra. This effect may be observed by looking through a handkerchief or umbrella top at a distant light. Brilliant diffraction effects are also obtained by looking at a source through a cobweb or feather, or from the light reflected from mother of pearl. In the latter case the effect is due to striations, as may be proved by transferring the effects to wax by pressure.

684. Diffraction of X-Rays.—For many years after their discovery the nature of X-rays was not understood, although it was suspected that they were very short waves similar to those of light.

The failure to detect reflection from plane mirrors or interference effects by gratings could be attributed to the impossibility of obtaining sufficient smoothness of surface or fineness and equality of spacing. It occurred to Laue in 1912 that the atoms of a crystal may be arranged with sufficient regularity to produce diffraction effects. The idea was verified by passing a narrow pencil of X-rays through a small hole in a lead screen and then through a crystal such as rock-salt. The X-rays then fell upon a photographic plate and produced a dark spot which was surrounded by a number of dark points arranged with regularity. The effect is similar to that obtained by looking at a distant light through a fine wire gauze or a cambric handkerchief, showing that the X-rays are due to wave motion. Bragg[1] succeeded in obtaining strong reflection effects from crystals due to the cumulative effect of many parallel layers of atoms. By Huygens' construction it may be proved that the maximum effect will be observed in that direction for which the angle of reflection is equal to the angle of incidence. The effect of a single layer is too small to be observed, but, if d is the distance between the atomic layers, i the angle of incidence, λ the wave length and N an integral number, when the further condition $2d \cos i = N\lambda$ is fulfilled, as in the case of the colors of thin plates, the reflections from many planes will be combined in the same phase. If the incident rays pass through a narrow slit, line spectra may be obtained in this way (§680). The distance d may be obtained from the density of the crystal and Avogadro's number (§313), and from this the wave-length may be calculated. Each element when used as a target gives a characteristic spectrum, with wave-lengths of the order of one thousandth to one ten-thousandth of the lengths of the waves in the visible spectrum. Similar effects with ordinary light are obtained in the reflection of white light from the laminæ of a chlorate of potassium crystal.

Recent work on X-ray spectra by Siegbahn and others has attained a precision comparable with that in the measurement of the wave-lengths of ordinary light. The resulting spectra are quite complex and are of great importance in modern theories of atomic structure (§731). Recently Bergen Davis and Siegbahn have obtained direct evidence of refraction of X-rays by passing them through prisms of various materials. The rays are bent *very slightly away* from the base of the prism, corresponding to an

[1] Bragg, X-rays and Crystal Structure.—Kaye, X-Rays.

index of refraction slightly *less* than unity. By using nearly grazing incidence A. H. Compton has obtained total reflection of X-rays from metal surfaces (similar to total reflection of ordinary light, §665), and by using an ordinary reflecting grating under such a condition he has measured the wave-length of X-rays in the manner employed for ordinary light. Hence X-rays can be reflected, refracted, and diffracted like ordinary light and the seemingly unusual phenomena associated with them are due solely to their extremely short wave-length. Finally, by knowing the wave-length of the X-rays employed, it is possible to calculate the value of d, the distance between atomic layers in the crystal used for reflecting the X-rays. This process, as carried out by Bragg and others, has led to the exact determination of the arrangement of atoms in the well-known types of crystals.

OPTICAL INSTRUMENTS AND MEASUREMENTS

685. The Eye is an essential part of any optical combination. Like a photographic camera, it is a closed chamber into which light can enter only through the lens. As the camera lens throws an image on a sensitive photographic plate which excites the silver grains, the lens of the eye forms a picture on the mat of sensitive nerve endings covering the retina. The amount of light entering the camera is regulated by an "iris" diaphragm of adjustable size; similarly the amount of light entering the eye is controlled by the size of the pupil, which automatically changes in diameter between the limits of about 2 and 5 mm. The parts of the eye are shown in Fig. 582. S is the sclerotic membrane, the outer enclosure of the eye. C is the cornea, a strong transparent membrane. I is the iris, the colored part of the eye, with a central orifice, the pupil, which admits light through the crystalline lens L, which focuses images on the retina R. The nerve endings covering the retina run together like the strands of a cable into the optic nerve O, which conveys stimuli to the brain. Muscles attached to the periphery of the lens can by their contraction or relaxation so change its focal length as to enable it to focus either distant or very near objects on the retina. This process is called accommodation. Two

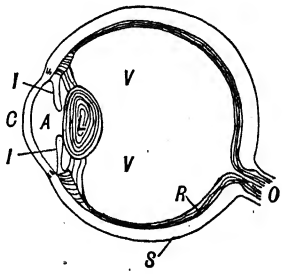


FIG. 582.

objects are clearly seen separately when the angle between them at the eye is a little less than 1′, or the distance between the retinal images is 0.005 mm. Details are, therefore, more clearly seen as an object is brought nearer, since the angle subtended by it and the size of the retinal image are then larger; but there is a limit to the power of accommodation of the eye, so that usually no object nearer than about 25 cm. can be seen clearly. This is called the distance of most distinct vision. The normal adjustment of the eye when at rest, is for "infinity," as may be verified by suddenly raising the eyes when they have been unemployed and looking toward distant objects. They will be in focus.

Between the cornea and the crystalline lens is the aqueous humor, *A*, and between the lens and the retina is the vitreous humor *V*, both transparent fluids with a mean index of refraction equal to 1.336. The lens is built up of transparent horny layers, increasing in density, hardness, and refractive power toward the center. The index of refraction of the outer layer is 1.405, and of the central region 1.454. The average index of refraction is about 1.437. This increase in density toward the axis serves partly to correct spherical aberration, which is also diminished by the iris diaphragm.

Objects such as printed letters can be seen clearly through a pin-hole in a card, even if they are as close as 2 cm. to the eye. This has been attributed to an over-correction of the lens for spherical aberration, so that a narrow pencil passing through the axis of the lens has a very short focus. It is obvious that so much over-correction would be worse than no correction at all. As a matter of fact, a pin-hole image is formed on the retina, the lens merely sharpening the effect. The fact that the apparent size of the object varies as the card is moved back and forth, the object remaining at rest, shows that the image is due mainly to the pin-hole.

686. Vision.—The retina is covered, except over the optic nerve, by a large number of very small fibrous bodies, "rods" and "cones," nerve endings which are in some way stimulated by light waves. Over the optic nerve is the "blind spot," so-called because if the image falls on this part of the retina it ceases to be visible. By closing one eye and looking steadily with the other at one of two small objects about two inches apart, a distance may be found at which the other object will disappear. Excitation of the optic nerve lasts about one-tenth of a second after the stimulus ceases, so that if intermittent stimuli are applied at intervals less than this a steady effect is produced. This is called *persistence of vision.*

The trail of the lighted end of a cigar if it be rapidly moved and the apparent continuity of moving pictures depend on this effect.

Sometimes the normal spheroidal shape of the lens is altered so that the curvatures are not the same in different planes. Light from a point will then pass through the eye as an astigmatic pencil with two focal lines instead of a point image (§661). Horizontal and vertical lines at the same distance cannot be simultaneously brought into focus. Such eyes are said to be astigmatic. Other defects arise from change of curvature or from loss of the power of accommodation. If eyes are short sighted, the principal focus falls short of the retina, and distant objects cannot be clearly seen. If eyes are long sighted, the principal focus is on or near the retina, and images of near objects cannot be formed on the retina. For the first defect concave spectacles are the remedy; for the second they must be convex.

In normal eyes the nerve endings on which fall corresponding points of the two retinal images lead to the same nerve centers, so that the two pictures are exactly superimposed and a more intense effect secured than with one eye alone. If one eye-ball be forcibly twisted out of position double images will be seen. A further advantage given by two eyes is that an object is viewed from two slightly different directions, which gives the impression of relief. This principle is applied to the stereoscope, in which two photographs taken from slightly different locations are viewed by each eye separately. The two images will be superimposed in such a manner that the object appears to stand out in space.

With two eyes it is also easier to estimate distances than with one. There is an angle between the two lines of sight to the object, which the brain unconsciously estimates. In general the sizes of objects are inferred from their angular magnitudes and estimates of their distance based on experience, or by comparison with adjacent objects, such as trees and houses, the sizes of which are approximately known. Such estimates are influenced by the clearness with which details are seen. In places where the atmosphere is unusually clear, as in Arizona, this leads to the underestimation of distance. Conversely, objects seen in a fog appear to be more distant than they are, owing to the indistinctness of their details. The angle subtended by them, however, corresponds to the actual distance, hence they loom larger than they are.

Irradiation is the apparent increase in the sizes of objects as they become brighter. The crescent of the new moon, for example, looks larger than the remainder of the disk, the "old moon," which is illuminated by the earth alone. The filament of an incandescent lamp appears to increase in size as it passes from ordinary temperatures through red and white heat. This effect was long supposed to be due to the spreading of the retinal image on account of stimulation of nerves outside of its boundaries, in much the same way that an over-exposed photographic image is affected. It is now believed by some that the effect is due merely to spherical aberration of the eye, which becomes more noticeable as the intensity of the sources increases.

687. The Simple Microscope or magnifying glass is a single convex lens through which objects at or within the principal focus of the lens are viewed. As shown in Fig. 583, an enlarged virtual image $A'B'$ is formed subtending at the lens the same angle α as the object AB. The linear size of this image is determined from the relation

$$I = (v/u)O$$

As the normal adjustment of the eye is for infinity, the object is usually placed at or very near the principal focus. In no case can the image be seen clearly when nearer than the limit of distinct vision. In the case of any optical instrument, the magnification depends not on the linear size of the image, but on the angle which it subtends at the eye, since the linear size of the retinal image depends on this angle. If the eye is very near the lens this angle is substantially that subtended from the lens. The lens simply increases the power of accommodation of the eye, so that the object may be brought nearer and thus subtend a greater angle. With the unaided eye, the greatest detail is observed at the distance of most distinct vision (25 cm.) where it subtends the angle β (Fig. 583). With the lens, the object is brought nearer, approximately to the principal focus, and the angle subtended by it increases from β to α. The magnification M of the retinal image is, therefore, α/β. If the distance of the object AB from O is u, and that of the image $A'B'$ from O is $v = d$ (the limit of distinct vision), then, as shown in Fig. 583,

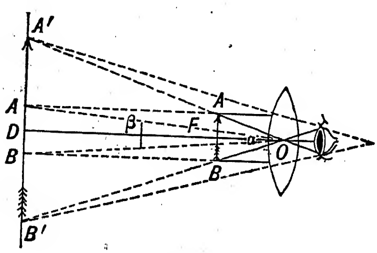


FIG. 583.

$$AB = 2d \tan \beta/2 = 2u \tan \alpha/2$$

Therefore, if these angles are small

$$M = \alpha/\beta = d/u$$

From this result and the lens formula it is readily seen that M is also equal to $1 + d/f$. Since in practice d is usually large compared with f, M is practically d/f.

688. Power of Lenses.—When the image is at infinity, the magnification is, as just shown, inversely proportional to the principal focal length, hence $1/f$ is a measure of its "power" (as opticians call it). The practical unit of lens power is that of a lens with a focal length of one meter. This unit is called a *diopter* or *dioptric.* The power of converging lenses is considered positive, that of diverging lenses negative. The relation deduced in §662 shows that the power of a number of lenses in contact is the algebraic sum of their individual powers.

689. Eye-pieces.—The part next the eye of an optical train of lenses, such as those of telescopes and compound microscopes, usually consists of some form of simple microscope known as an

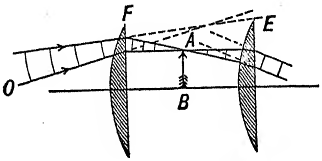


FIG. 584.

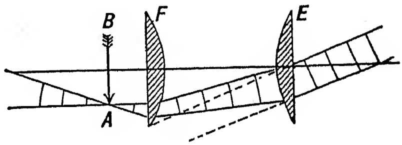


FIG. 585.

eye-piece. With a single lens, much of the light from the real image formed by the objective O, which is usually viewed through the eye-piece, would be lost. In order to avoid this, light is gathered in toward the axis by a second lens, called the field lens F (Fig. 584). Nearly all the light would pass by the edge of the eye lens E if F were absent. It may be shown that a combination of two lenses of the same kind of glass is nearly achromatic if they are placed at a distance from each other $d = (f_1 + f_2)/2$. This property is utilized in most eye-pieces which consist of a field lens and eye lens.

In Huygens' eye-piece the field lens F has a focal length $f_1 = 3f_2$ (Fig. 584). Hence $d = 2f_2$ and if the image due to the objective and the first lens is formed half way between the lenses the emergent light will be parallel and a virtual image formed at infinity. The lenses are convex toward the incident light and of such curvature as to reduce the spherical aberration to a minimum. This type of eye-piece is generally used in microscopes.

In the Ramsden eye-piece (Fig. 585) $f_1 = f_2$. If the lenses were placed apart at the distance $(f_1 + f_2)/2$, dust particles on the field lens would be visible through the second. In order to avoid this, the lenses are usually placed at a distance of $2f/3$. The principal

focal point of the combination is at a distance $f/4$ in front of the first lens. The object, or the real image due to the objective, is at this point, and the final virtual image is at infinity. The chromatic aberration is small, and the spherical aberration is reduced by using plano-convex lenses with convex surfaces facing each other. The Ramsden eye-piece is used in a surveyor's transit and in other instruments where a cross-hair is needed, the cross-hair being placed at AB.

In all these eye-pieces the emergent red and violet rays are nearly parallel, hence the virtual images formed by the different colors subtend very nearly the same angle at the eye, and are, therefore, of the same size, but the focus on the retina is not equally sharp for the different colors.

690. Compound Microscope.—In order to extend the limit of magnification beyond the point obtainable with a simple micro-

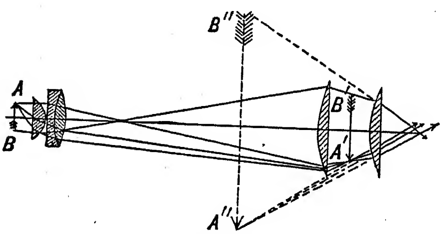


FIG. 586.

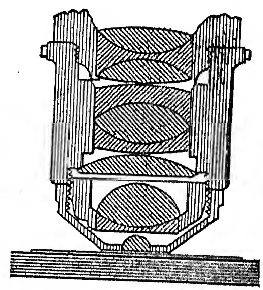


FIG. 587.

scope, a combination of lenses is used. An enlarged real image $A'B'$ (Fig. 586) is formed by an object lens or train of lenses, and this image is further enlarged by an eye-piece, such as that of Huygens, used as a simple microscope, which gives a virtual image $A''B''$. The front lens of the objective train is usually of the hemispherical form described in §659, which has great angular aperture, with very little spherical aberration. There are in addition a number of other lenses of different shapes and kinds of glass, so combined as to reduce spherical and chromatic aberration to a minimum and to give a plane focal surface. A typical combination is shown in Fig. 587.

The magnification due to the objective, of focal length f_1, is

$$M_1 = I_1/O = v_1/u_1 = v_1/f_1, \text{ approximately.}$$

That of the eye-piece is approximately, as shown in §687,

$$M_2 = I_2/I_1 = D/f_2$$

where D is the minimum distance of distinct vision. The magnification due to the combination is

$$M = M_1M_2 = I_2/O = LD/f_1f_2 \text{ approximately,}$$

where L is the distance between the objective and the eye-piece.

The minimum distance between two small objects A and B seen through a microscope which will permit of clear separation of their diffraction images is obtained by a slight modification of the expression found for the minimum angle of resolution, $\alpha = \lambda/AB$(§682). The minimum value of this distance is thus found to be $\lambda/2$. Since this is proportional to the wave-length, details which may be clearly seen when the object is illuminated by blue light will be indistinct when red light is used.

691. Astronomical Telescope.—The object glass of a telescope forms a real and of course greatly reduced image $A'B'$ of a distant

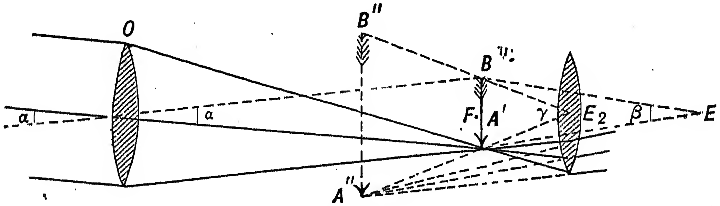


FIG. 588.

object (Fig. 588). The object and its image subtend the same angle α at the objective, and the object subtends practically the same angle α if viewed directly by the eye. If, however, the eye views the image formed by the objective at the distance of most distinct vision, (from the point E), this image will subtend an angle β which is larger than α, and the apparent magnification is $M_1 = \beta/\alpha$. When this image is viewed through an eye-piece (the eye now being at E_2), there is further magnification, the image subtending the larger angle γ. The magnification due to the combination is

$$M = M_1M_2 = \frac{\beta}{\alpha} \times \frac{\gamma}{\beta} = \frac{\gamma}{\alpha} = \frac{f_1}{f_2}$$

where f_1 is the focal length of the objective, and f_2 of the eye-piece. In a telescope the final image is usually at infinity and this formula

for M is then accurate, except for the fact that the arc has been substituted for the chord, in the case of each angle (compare end of §687).

The limiting angle of resolution between two linear sources is proportional to $1.22\lambda/A$, where A is the diameter of the objective (§682).

For astronomical purposes there is no disadvantage arising from the fact that an inverted image is formed by a telescope, but when the instrument is to be used for terrestrial purposes it is necessary to add an additional lens or pair of lenses to reinvert the real image formed by the objective. This adds inconveniently to the length of the tube. If the image is inverted by reflection from a combination of prisms the length may be diminished, but for most purposes where only small magnification is required the form of telescope devised by Galileo is most convenient.

692. Dutch or Gallilean Telescope.—This type is used for opera glasses and for marine glasses. As shown in Fig. 589 an erect virtual

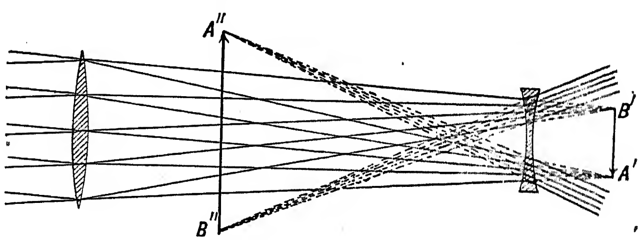


Fig. 589.

image is formed, the magnification being $M = f_1/f_2$ (§691). The tube has a length approximately equal to the difference between the focal lengths of the objective and the eye-piece, while in the ordinary telescope the length is the sum of these distances.

693. Reflecting Telescope.—The objective lens may be replaced by a large concave mirror. In this way chromatic aberration may be entirely avoided, but spherical aberration is more troublesome than with refractors. There is no aberration however, if the mirror is a paraboloid and the object is at infinity. As the real image is formed along the axis of the mirror and in the path of the incident light, special devices are necessary in order to view it. In the Newtonian telescope the image is reflected to one side by a small right-angled prism which cuts off very little light, and is viewed by an eye-piece in the side of the tube. Herschel tipped the mirror

slightly so that the image was formed at the edge of the open end of the tube, at which point the eye-piece was fixed. In other forms a small mirror in the axis reflects the image back into an eye-piece set in the center of the objective itself, so that it can be viewed from behind. The largest telescope of this type is that at Mount Wilson, California, with a mirror 100 inches in diameter. Work is now in progress on a 200 inch mirror.

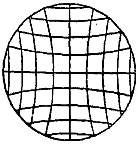


FIG. 590.

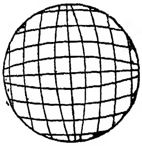


FIG. 591.

694. Photographic Camera.—This is an instrument in which the image formed by a lens falls on a sensitive photographic plate. The requirements demanded for the lens are exacting and in some cases contradictory to each other. It must give images free from spherical and chromatic aberration, and in many cases it must have great light power and a large field of view. The focal surface must be plane, and the magnification must be the same in all parts of this plane, so that no distortion is produced. The depth of focus must be great, that is, objects at different distances must have images approximately in focus at the same time on the plate. As the film is most sensitive for the shortest waves, the lenses must be corrected for the violet and the yellow, instead of blue and red. A diaphragm with a small opening is used in front of the lens, if it is a single achromatic combination, such as is used for landscape work. This reduces spherical aberration and at the same time gives a greater depth of focus (approximating to the principle of the pinhole camera, in which the focus is independent of the distance). A diaphragm with a single lens results in a distorted image, however, as shown in Fig. 590, which represents the distortion of the image of a quadrilateral network with the diaphragm in front of the lens, and Fig. 591, which gives the effect due to a diaphragm behind the lens. The cause can be shown to be due to the difference in deviation of pencils passing through the center and the edge of the lens respectively. If two lenses are used, with the diaphragm at the optical center of the combination, these distortions correct each other.

As shown in Fig. 592, the perfect symmetry of the incident and the transmitted secondary axes AA', BB', CC', etc., with respect to the opening O shows that the distances AB, BC, etc., in the object are in the same ratio as the corresponding distances $A'B'$, $B'C'$, etc., in the image, so that there is no

distortion, and if A, B, C are in the same plane, $A'B'C'$, etc., must be in the same plane. Such lenses are called *rectilinear or orthoscopic doublets.*

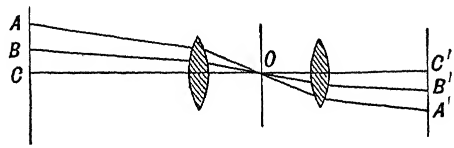


FIG 592

The size of the photographic image of a distant object is nearly proportional to the focal length of the lens. It is, however, inconvenient to give a great length to the camera box This difficulty is avoided by the use of the *teleobjective,* in which a concave lens L_2 is placed behind the converging lens L (Fig. 593). The divergent effect of this lens gives a virtual focal length equal to PF, while the camera box has the smaller length LF. By suitably adjusting the focal lengths and separation of the two lenses PF can be made many times as great as LF. A greatly enlarged image is thus secured, but the field of view is correspondingly reduced

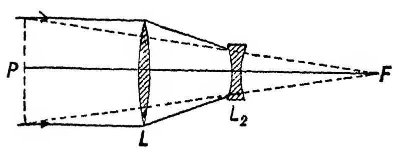


FIG 593

695. The Projection Lantern is used to throw an enlarged image $A'B'$ of more or less transparent objects on a screen The object AB (Fig 594) is illuminated by a source S, through a condenser C, consisting of two thick planoconvex lenses, with convex sides facing each other. The source is usually the electric arc or incandescent lamp. The focusing lens L is generally of the photographic doublet type, in order that an undistorted image may be formed on the screen. The object of the condenser is not only to illuminate the object, but also to enlarge the field beyond the limit which would otherwise be set by the cross-section of the focusing lens

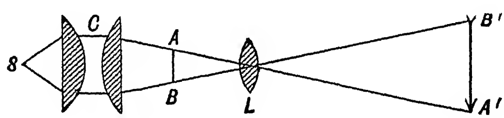


FIG 594.

696. The Range-finder is an instrument used chiefly in naval and military operations to determine the distance of objects. Two lenses are mounted at the ends of a tube several feet long, which is the base of a triangle of which the object is the vertex and the two lines of vision the sides. The two images formed by the lenses are reflected by prisms into an observing telescope fixed to the middle of the tube From the displacement of the images the angle at the vertex may be determined, and with this and the known length of the base the distance of the object may be calculated or read from an empirical scale.

697. The Spectroscope (Fig. 595) is an instrument for analyzing complex radiations by prismatic dispersion (§648) or by the diffraction grating. In order to secure as complete separation of the colors as possible, or a "pure" spectrum, a narrow slit must be used as a source, so that the colored images of the slit will overlap as little as possible. These images, as in all optical instruments, are really diffraction patterns, and it is therefore also necessary to use large apertures for the lenses and prism or grating, in order to make the images, or spectral "lines" as sharp as possible (§683). The larger the dispersion the greater the separation of the images For given dispersion, the length of the spectrum is proportional to the focal length of the observing telescope, but this merely affects

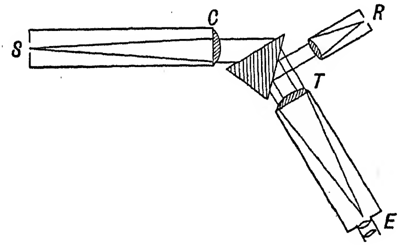


FIG. 595.

the scale of the spectrum, not the resolution of the lines or the clearness of detail.

The plan of the ordinary form of spectroscope is shown in Fig. 595. The essential parts are: a narrow slit, S; a collimating lens C, which converts the wedge of light from the slit into a parallel beam; a prism to disperse the colors; a telescope lens T, with which real images of the slit are produced at the focus of the eyepiece E. If light of an infinite number of colors is emitted by the source, the infinite number of partially overlapping images forms a *continuous* spectrum; if only a finite number of colors is emitted, there will be a finite number of slit images, giving a discontinuous or *line* spectrum. A homocentric pencil (cone) incident on a prism remains homocentric after transmission at the angle of minimum deviation; for all other angles of transmission, it becomes astigmatic, and no true image is produced. The condition of homocentricity cannot be fulfilled for all colors simultaneously except when the incident light is parallel, in which case light of each color emerges in a parallel beam. For this reason the collimator is necessary. This lens must

be achromatic. So far as purity of spectrum is concerned, it is evidently unnecessary for the telescope lens to be achromatic, but it is usually corrected, in order that all the colors may be at once in the focus of the eye-piece. Positions in the spectrum may be referred to the image of a scale R reflected from the side of the prism.

This instrument is called a spectrograph when the telescope is replaced by a camera for photographing the spectrum, and a spectrometer when provided with a graduated circle for measuring the angular deviation of the light. The direct-vision spectroscope, usually made in small sizes for pocket use, has a combination of crown and flint prisms, as shown in Fig. 596. The mean deviation is zero, but there is some residual dispersion which gives a short spectrum (§652).

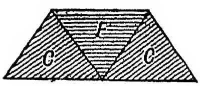


Fig. 596.

A plane diffraction grating may replace the prism of a spectroscope, or spectrometer. With the latter the angular deviations of the diffraction maxima may be measured and the wave-lengths determined by the relation deduced in §683.

698. The Concave Grating was Rowland's greatest contribution to spectroscopy. The lines are ruled at equal distances on the surface of a concave mirror of speculum metal which focuses as well as diffracts the light. If R is the radius of curvature of the mirror (Fig. 597) and if the slit is at any point on the circumference of a circle having the radius as its diameter, it is found that the spectra of all orders are in focus, along the circumference of this same circle, so that no lenses are necessary. Usually the grating G is mounted at one end and the eye-piece, or camera, E at the other end of a beam R equal in length to the radius of the grating. This beam has a swivel truck under each end, which travels on tracks at right angles to each other, with the slit at the intersection S. The variable distance SE between the slit and the eye-piece is proportional to $a \sin \alpha = N\lambda$ and therefore is proportional to the wave-length of the light focussed at E. If a curved photographic plate is placed at E perpendicular to R, it may be proved that the spectral lines on this plate form very nearly a normal spectrum (§683).

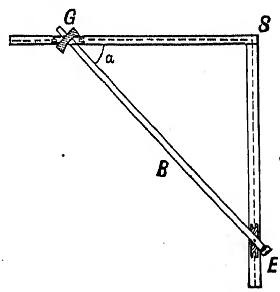


Fig. 597.

699. Michelson's Interferometer.—The surface of the glass plate P_1 (Fig. 598) is "half silvered," that is, the silver film is of such thickness that about one-half of the incident light is reflected. Light from the point S of an extended source falls on this surface

at A and is in part transmitted to the mirror M_1, in part reflected to the mirror M_2. From these mirrors it will be reflected, retrace its course, and some will finally reach the eye at E. If M_1 and M_2 are at the same optical distance from S, and if each is perpendicular to the rays that fall on it the light will appear to come from two exactly superimposed images of the source and there will be no interference. The plate P_2 is introduced merely to give the ray SAM_1 the same path in glass as the ray SAM_2, so that the optical and the geometrical paths will be the same. Now if monochromatic light of wave-length λ be used and if M_1 be displaced a distance $\lambda/4$, the waves that reach the eye will have traversed paths that differ by $\lambda/2$ and will destroy one another in the center of the field of view. A further displacement of $\lambda/4$ will restore the light. Thus by slowly displacing M_1, and counting the number of times, N, that the light reaches a maximum, the distance, d, through which M_1 has been displaced may be found from the relation

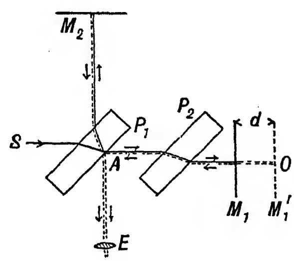


FIG. 598.

$$d = N\lambda/2.$$

Michelson has by this instrument measured the length of the standard meter of Paris in terms of wave-lengths of several of the spectral lines of cadmium, with an accuracy of about one part in two million. The wave-lengths of such lines are probably the most permanent and unchangeable standards of length which can be obtained. The interferometer has been used to make numerous other measurements of great accuracy. For example, the thickness of the thinnest water films has been found by means of it. If a film of thickness d be introduced in the path of one of the rays the change of optical path (§654) will be $2(n - 1)d$. This distance being measured by the interferometer and n being known, d is deduced.[1]

The above account refers to what is observed in the center of the field of view, the mirrors being adjusted so that the image of M_1 in the plate P_1 is parallel to M_2. Around the central point there are circular fringes which change in diameter as M_1 is moved. If M_2 and the image of M_1 are not quite parallel the effects will be similar to those due to a wedge-shaped film (§671).

[1] See Michelson, Light Waves and Their Uses.

The fringes will still be approximately circular but not concentric, the centers being on a line perpendicular to the edge of the wedge. As M_1 moves the fringes will sweep past any point in the field of view and it is this succession of fringes that is usually counted in the use of the instrument.

In the Fabry and Perot interferometer two thinly-silvered mirrors are mounted parallel, and a number of multiply-reflected pencils interfere with each other, producing a sharper system of rings. These are the Haidinger fringes (§674) of a thick "air plate," observed in the transmitted light. As in the case of a grating, the resolving power is larger when a great number of pencils are caused to interfere (§683).

700. Lummer-Brodhun Photometer.—A cube C is made of two right-angled prisms, as shown in Fig. 599. The hypothenuse surface of the one prism is plane, that of the other convex, with the vertex ground flat. These two surfaces are in close contact. The sources to be compared, S_1 and S_2, are mounted on an optical bench, over the center of which is the white screen W of paper or gypsum. The diffuse illumination from this screen is reflected from the mirrors M_1 and M_2 through the prism faces AB and BD. Light from S_1 and S_2 is transmitted without loss through the area of contact of the two prisms, and is totally reflected from the air film between those parts of the hypothenuse surfaces which are not in contact. If a telescope is focused on the region of contact through the side CD, light will enter it from S_1 by transmission and from S_2 by reflection. The field will be uniformly illuminated if the two sides of the screen W are equally illuminated; otherwise the area of contact will appear brighter or darker than the surrounding part of the field.

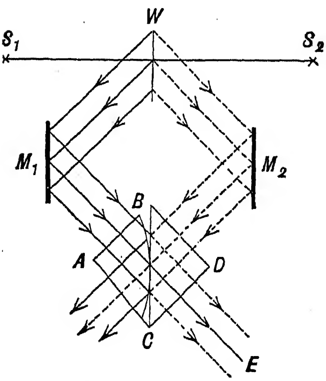


Fig. 599.

EMISSION AND ABSORPTION OF RADIANT ENERGY

701. Analysis of Radiation.—The methods by which radiation may be analyzed by the dispersion of colors, or waves of different length, have already been described (§697). The radiation from all known sources is complex—that is to say, it contains waves of more than one frequency of vibration. The principal types of emission spectra may be observed side by side if the image of a long electric arc is focused on a slit beyond which a prism and lens are placed so that a large spectrum is thrown on the screen. The light coming from the positive carbon forms a brilliant continuous spectrum including all the colors. Next to it is the discontinuous spectrum of the arc proper, the luminous flame, which contains

the vapors of carbon, various compounds of carbon, and any metals that may be present as impurities in the electrodes. This spectrum consists of a number of narrow lines due to the metals present, and several groups of bands, each composed of a large number of fine lines so spaced as to produce the effect of the shading of fluted columns in a line drawing—hence they are often referred to as fluted bands (see Fig. 603*a*). These appear to be due to the vapors of carbon (in the form of the molecule C_2) and to the compounds of carbon. The bands are especially strong in the violet, and this gives the arc its characteristic violet color. The violet bands are now known to be due to cyanogen (CN). Next to this is the spectrum of the negative electrode, which is at a much lower temperature than the positive and in which there is very little blue or violet.

As illustrated by the arc spectrum, there are two general types of *emission* spectra, the *continuous* and *discontinuous*, and the latter in turn may be divided into *line* and *band* spectra (§§704, 705).

702. Invisible Radiations.—It was found by William Herschel in 1800 that if a sensitive thermometer is placed in any part of the spectrum of the sun it will show a rise of temperature, this effect increasing in going from violet to red. It does not, however, cease abruptly at the boundary of the visible spectrum, but increases to some distance beyond it, and then gradually diminishes, the observed limit in any case depending on the sensitiveness of the thermometer. Evidently there is radiation which is less refracted than the red, and which by analogy we may conclude has waves of greater length than those of red light. It was shown by Herschel that this "radiant heat" is subject to the same laws of reflection and refraction as light, but their identity was not generally accepted until nearly fifty years later, when it was shown that the invisible radiation is capable of producing interference effects, and that it is likewise capable of dispersion and polarization. As the ideas in regard to the nature of heat crystallized it was seen that heat can be associated only with matter, but that the energy of this heat may be partly transformed into the energy of ether waves, and this, if absorbed by matter, will again appear as heat. It thus becomes clear how invisible radiations from a hot body can pass through an ice lens without melting it, and set fire to a piece of paper at the focus. The name *infra-red* is applied to these long-wave radiations.

The existence of *ultra-violet* radiations in the solar spectrum, with waves shorter than those of violet, is shown by means of the chemical effect produced on chloride of silver. Photographic films are very sensitive to the ultra-violet radiation, which is especially active in its chemical effects. It also excites strong fluorescence (§727) in many substances. If a strip of paper moistened in acidulated sulphate of quinine solution is held in the arc spectrum, the excited fluorescent light shows the existence of ultra-violet radiation in the spectra of both the positive carbon and the arc proper.

The non-visibility of the infra-red and ultra-violet radiations is due merely to the limitations of the eye. The eye will "resonate" to vibrations between certain limits of frequency, the photographic film or fluorescent screen to certain others; but if the receiving surface is blackened the energy of waves of all frequencies is almost completely absorbed, and by the amount of heat developed we may determine the amount of energy in any part of the spectrum.

703. Methods of Detecting Invisible Radiation.—Photography is a thoroughly satisfactory method of detecting even the shortest ultra-violet radiations so far discovered. This method cannot be used however, at the opposite end of the spectrum, as it seems impossible to make any photographic film which is sensitive to the major portion of the infra-red—it is difficult to make one which will even reach that portion lying just beyond the visible red. For this reason other less satisfactory methods must be employed, which are usually based on the heating effects produced. For one of the earliest instruments for the detection of infra-red radiation, the thermopile, as well as other more recent modifications, see §§332, 464.

704. Continuous Spectra.—It is a familiar experience that as the temperature of a body rises it first reaches a dull red heat, then yellow, and finally a dazzling white. Conversely, if the spectrum of the positive electrode of an arc light is thrown on a screen, and if the current is suddenly cut off, it will be observed that, as the carbon cools, violet, blue, green, and yellow disappear in succession, and finally the red. If a sensitive thermopile is placed far in the infra-red it will be found that appreciable radiation is still emitted long after the luminosity has disappeared.

Draper (1847) found that all bodies begin to glow at about the same temperature. The actual temperature in any case depends somewhat on the sensitiveness of the eye, but is not far from 400°C. *Draper's law* is approximately true for all colors and temperatures—that is to say, all solids begin to radiate a perceptible amount of red,

or yellow, or violet, or any particular "heat color" in the infra-red at the same temperature. The spectral distribution of energy may be shown by plotting a curve with wave-lengths as abscissæ and ordinates proportional to the galvanometer deflections observed as the thermopile passes through the spectrum. Fig. 600 shows a series of such curves for temperatures ranging from 836° to 1377° absolute, the source consisting of a strip of blackened metal elec-

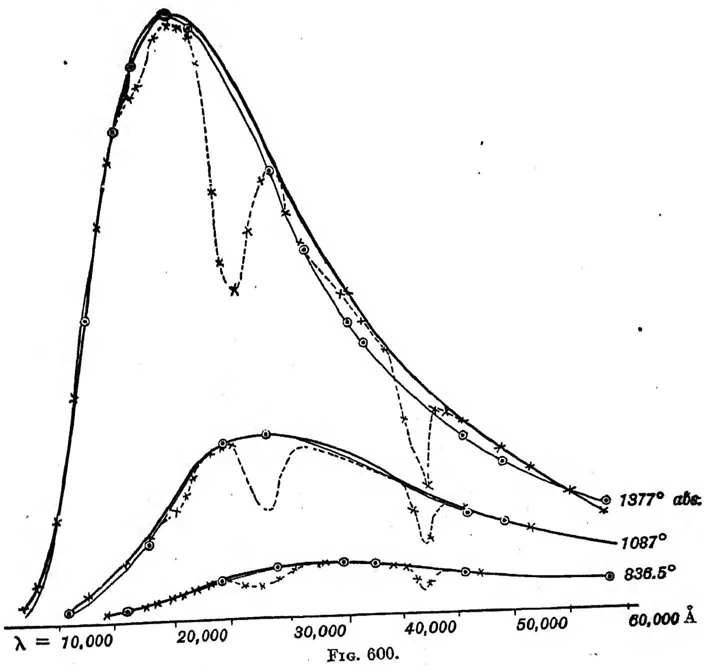


FIG. 600.

trically heated. The depressions in the curves are due to absorption by carbon dioxide and water vapor. The general character is the same for all solids, but differences in the ordinates may arise from differences in the state of the surface, whether black, or rough, or polished, etc. *A "black body" is defined, technically, as one which absorbs completely all wave-lengths falling upon it.* Investigation shows that there is a very definite relation between the absolute temperature of the source if it is black, or approximately so, and the wave length corresponding to the maximum ordinate of the energy curve, such that $\lambda_m T$ = constant (see §705). The best

value of this constant, for a black body, is 2884, if the unit of wave-length is the micron, μ, or 0.001 mm.

The total energy emitted by an incandescent source is proportional to the area included between the energy curve and the axis of wave-length, λ. That part of the energy which produces luminosity is included between the limits of the visible spectrum. The ratio of the latter to the former area might be thought to give the luminous efficiency of the source. This is not the case, however, since it neglects the varying sensitiveness of the eye to various colors. When this latter factor is also considered, one obtains the true "luminous efficiency," which, when expressed as a percentage, becomes the "reduced luminous efficiency." Fig. 601 illustrates the spectral energy curves of the positive pole of an arc light, of an incandescent light, and of a piece of red hot carbon. The energy lying in the visible region is given by the shaded area under each curve. In the highest curve, that for the arc, this is about 10 per cent of the whole area. The reduced luminous efficiency of the arc is, however, somewhat less than 2 per cent because the sensitivity of the eye is very low for the wave-lengths of visible light lying near the limits V and R. This sensitivity reaches a maximum at the wave-length 0.555μ in the yellow-green, and in order to have a reduced luminous efficiency of 100 per cent, a source would have to emit radiation of this wave-length only. Luminous gases, which radiate "selectively" in the visible, have a higher efficiency than incandescent solids. As will be seen from the curves of Fig. 601, the efficiency of the latter increases rapidly with temperature. In the case of the gas-filled tungsten lamp, it is possible to use a temperature as high as 3290°K., but even this yields a reduced luminous efficiency of only 4.4 per cent.

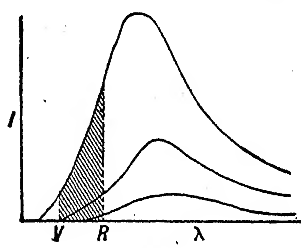


Fig. 601.

705. Law of Radiation.—It may be assumed that electromagnetic waves are set up by agitation of the electrons, or atoms in the case of the long waves, within the molecules of matter, and that the frequencies of the waves are dependent upon the energy of the molecules. In solids the molecules are so close that there can be little chance for them to radiate without constraint, with their natural periods. Owing to frequent collisions, a wide range of velocities and vibration frequencies must exist. The greater part of the radiant energy will be due to the large number of molecules which have energies corresponding to

the mean velocity. There will be relatively few molecules which will have extreme energies, either large or small, and therefore the longest and shortest of the waves will have relatively a small amount of energy. As the temperature rises there will be a general increase of kinetic energy, many molecules moving faster but none slower than before, so that both the maximum energy and the maximum rate of vibration of the excited ether waves will move toward the violet end of the spectrum. The source will rise to red and finally to white heat. It is thus evident that a spectral intensity curve must be a sort of probability curve based on the distribution of velocities of the molecules with its ordinates exaggerated on the side of the violet, as illustrated in the curves of Fig. 600.

By such reasoning Wien theoretically deduced the relation $\lambda_m T$ = constant, known as Wien's law. Later Planck established a general relation between the intensity of radiation corresponding to a given wave-length λ, or a given frequency ν, and the absolute temperature of the source, as follows:

$$I_\lambda = \frac{c_1}{\lambda^5}\left(e^{\frac{c_2}{\lambda T}} - 1\right)^{-1} \text{ or } I_\nu = \frac{h\nu^3}{V^2}\left(e^{\frac{h\nu}{kT}} - 1\right)^{-1}$$

where e is the base of the natural system of logarithms, T the absolute temperature, V the velocity of light, k = the gas constant R divided by Avogadro's number, and h a universal constant $= 6.55 \times 10^{-27}$ (the Planck constant). This relation holds within wide limits for bodies which are black or approximately so. For such bodies the value of $c_2 = hV/k = 1.432$, $c_1 = 2\pi hV = 3.697 \times 10^{-5}$. The law $\lambda_m T$ = constant $= c_2/4.9651$ may be deduced by differentiating the above expression for a maximum value of I_λ, and Stefan's law by integrating the intensity over the whole spectrum (see §337).

The expression for I_ν was derived from the assumption that electric oscillations in the molecules do not emit energy continuously, but in definite "quanta," a quantum being equal to the constant h multiplied by the frequency ν. This was the origin of the "quantum" theory of radiation, which seems to have no rational basis of mechanical or electro-magnetic theory, but has been wonderfully successful in describing the phenomena of radiation (see §757).

706. Discontinuous Spectra were noted by a number of observers during the first half of the nineteenth century, and it was at least dimly realized that in some cases the appearance of the spectrum is characteristic of the substances emitting the radiation. There are two types of discontinuous spectra, known respectively as *line* and *band* spectra.

In 1860 Kirchhoff and Bunsen established definitely the law that all gases and vapors give discontinuous spectra, and that these spectra are perfectly characteristic of the substance. They discovered rubidium and cæsium by the application of this principle, which has been fruitful in the discovery of other new elements, notably helium and the rare atmospheric gases in recent times.

At first only the spectra of flames were studied, but later it was found that the electric spark between metallic terminals gives lines due both to the electrodes and to the surrounding atmosphere, and that if the electric discharge passes through a gas in a partially exhausted tube the luminosity is confined to the gas, and the metallic lines disappear. Only in exceptional cases is it possible to make a gas luminous except by the electric discharge.

707. Line Spectra are given by the metals and salts of the sodium and calcium groups in the Bunsen flame, and also by a number of other metals if spray from solutions of their salts, or ions caused by the electric spark, are passed into the flame. The spectrum of the electric spark or arc between electrodes composed of or coated

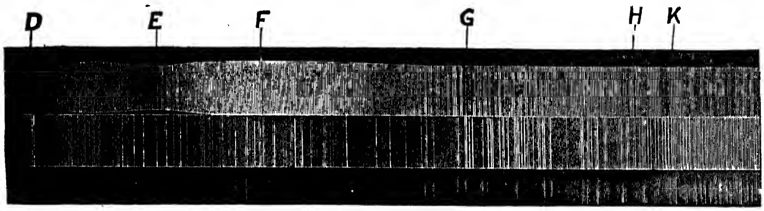


FIG. 602.

with any metals or their salts contains many more lines than that of the flame. The number may be very large, ranging from a dozen or so in the case of the alkali metals to many thousands, in the cases of iron and uranium.

It seems that no two elements have any common lines, but the spectrum of a given element may show differences in the number and the appearance of the lines according to the nature of the source, whether flame, arc, or spark. The middle part of Fig. 602 shows the arc spectrum of iron, the lowest part is the spark spectrum of the same metal, the highest part is the solar spectrum, which shows many absorption lines coinciding with the emission lines of iron. All salts of the same metal give the same line spectrum, although in some cases they give band spectra as well, which may be characteristic of the salt (§708). The lines of the non-metallic components do not appear with those of the metal except in rare cases. Intense electric discharges through a non-metallic gas at ordinary pressures or in vacuum tubes give line or band spectra.

708. Band Spectra are usually composed of fine lines, as shown in Fig. 603a (the spectrum of part of the carbon arc, due to cyanogen).

The light of the green cone in a Bunsen flame gives a very similar spectrum, due to the carbon molecule C_2. The salts of the calcium group of metals have flame spectra containing both lines and bands. Thus all salts of calcium give the same flame spectrum under ordinary conditions, but if calcium chloride, for instance, is placed in a flame supplied with hydrochloric acid, an entirely different band spectrum is produced, and still another if the flame is supplied with calcium bromide and an excess of hydrobromic acid. The inference is that in these cases the bands represent the characteristic

FIG. 603*a*.

spectra of the compounds, and that the spectrum observed under ordinary conditions is that of the oxide, due to reaction with atmospheric oxygen.

Nitrogen gives a band spectrum very similar to that of cyanogen if a feeble discharge passes through it, but an entirely different line spectrum if the discharge is very intense (Fig. 603*b*). Nitric oxide gives a characteristic band spectrum in the ultra-violet similar to that of nitrogen. All the compounds of mercury with chlorine, bromine, or iodine, give characteristic band spectra with feeble discharges, and the line spectrum of mercury with strong discharges.

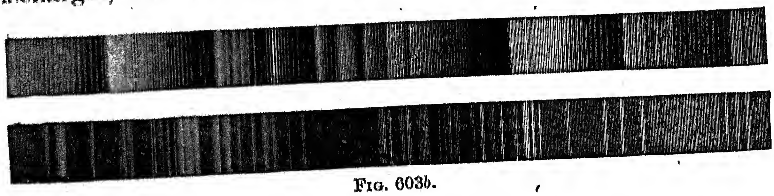


FIG. 603*b*.

The same is true in a number of other cases. All these facts are consistent with the view that band spectra are characteristic of the molecular state of either elements or compounds. Intense discharges, by dissociating the molecule, will produce line spectra, characteristic of the atomic state. The salts of the alkali metals are so easily dissociated that they give only line spectra in the flame.

709. Limits of the Spectrum.—The very short ultra-violet waves are absorbed by all gases, and by most lenses and prisms. With fluorite prism and lenses in a vacuum Schumann reached a limit of about .000012 cm. and later Lyman, Millikan, and recently

Edlén and Ericson with concave gratings have obtained much shorter wave-lengths. The ordinarily used unit of wave-length is the Ångström unit, equal to 10^{-10} meter, and hence sometimes called a tenth-meter. Other units frequently used are the micron $\mu = 10^{-6}$ meter $= 10^{-3}$ mm, and the millimicron $m\mu = 10^{-9}$ meter. Expressed in these units, some wave-lengths are given below. Most substances are opaque to very long waves, and some of the longest waves mentioned were obtained by the method of selective reflection described in §712, the wave-length being then measured by a coarse grating.

It has been proved that X-rays (§542) and gamma rays (§549) are similar to light waves, but very much shorter. Their lengths differ, but the order of their magnitude is given in the table below.

	Ångström Units	μ
Gamma rays	0.1	0.00001
X-rays	1	0.0001
Shortest ultra-violet waves (Edlén and Ericson)	76	0.0076
Shortest visible waves (violet), about	3,800	0.38
Violet, about	4,000	0.4
Blue	4,500	0.45
Green	5,200	0.52
Yellow	5,700	0.57
Red	6,500	0.65
Longest visible waves (red)	7,700	0.77
Longest waves in solar spectrum, more than	53,000	5.3
Longest waves transmitted by fluorite	95,000	9.5
Longest waves by selective reflection from rock salt	500,000	50.0
By reflection from potassium chloride	612,000	61.2
Longest waves from mercury lamp (Rubens)	3,250,000	325
Shortest electric waves (Nichols and Tear)	2,200,000	220

710. General Absorption.—When radiation falls on matter a portion is reflected, another absorbed, and if the substance is transparent or very thin a part is transmitted. Black substances, such as lampblack and copper oxide, reflect and transmit very little, the absorption being almost complete. Most substances black to visible radiation are also black to the ultraviolet and infra-red waves, but there may be exceptions—for example, a sheet of hard black rubber is opaque to visible radiation, but transparent to waves beyond the red. Substances like that last mentioned, which absorb certain radiations and transmit others, are said to exercise selective absorption.

711. Selective absorption is characteristic of nearly all substances. Familiar examples are red glass, which transmits red and some infra-red but no other visible colors; blue cobalt glass, which transmits blue and violet and a little red and green in narrow regions; green, which transmits almost all the colors, but a larger proportion of green; chlorophyll solution, potassium permanganate, the aniline colors, and solutions of the rare earths, didymium, etc. In most cases the absorption bands are wide and diffuse; in the case of the rare earths they are almost as narrow as spectral lines, so that the solutions appear almost colorless, no large amount of any one color being absorbed; the vapors of iodine, nitrogen peroxide, and some other substances have fluted absorption bands, grouped somewhat like the lines in the nitrogen bands and lying in the visible spectrum. Transparent gases, such as hydrogen, nitrogen, oxygen, etc. have band absorption spectra lying wholly in the far ultra-violet. Many solid substances such as glass, quartz and rock salt are very transparent within wide limits, beyond which they are completely opaque. Color screens of stained gelatine films are now made to give almost monochromatic light.

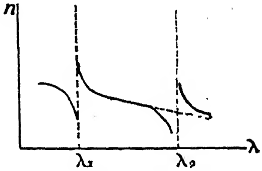


FIG. 604a.

FIG. 604b.

Ordinary glass is opaque to waves shorter than about 3500 Ångström units, and longer than about 25,000 Ångström units. Quartz is transparent between the wave-lengths 1800 and 300,000, and for some longer waves. Rock salt is transparent between 1800 and 180,000, and fluorite, one of the most transparent substances, will transmit ultra-violet waves from about $\lambda = 1250$ to $\lambda = 95{,}000$. All of these materials are more or less transparent to very short waves, in the X-ray region.

712. Anomalous Dispersion.—Substances which possess strong regions of selective absorption in the visible spectrum exhibit an irregularity in the curve of index of refraction against wave-length which gives rise to anomalous dispersion. Thus iodine vapor transmits only the red and violet, and *the red is refracted more than the violet.* This phenomenon is shown by a large number of substances, such as aniline dyes and the vapors of the alkali metals. For these, the curve of normal dispersion (§651) no longer holds through the visible region. The index of refraction is abnormally increased on the red side of the absorption band and diminished on the violet side. The dispersion curve of a substance having absorption bands at two wave-lengths λ_0 and λ_1 is shown in Fig. 604*a*. Such a curve may be shown directly by passing light successively through a glass prism which produces a spectrum in the horizontal direction, and through a prism of sodium vapor arranged to deviate the light vertically. Fig. 604*b*, which was photographed in this way, illustrates the anomalous dispersion of sodium vapor in the neighborhood of the *D* lines.

Between the wave-lengths λ_0 and λ_1, the curve of Fig. 604*a* resembles the normal dispersion curves shown in Fig. 539*b*. This suggests that the so-called "normal" curve is obtained when the absorption bands of the material lie outside of the visible region, and that all substances would show "anomalous" dispersion if the investigations were extended over a sufficient range of wave-lengths. In fact, we find that substances like glass and quartz have strong absorption bands in the ultra-violet, and others in the infra-red.

The connection between absorption and anomalous dispersion is now explained by the interaction between the charged particles (electrons, ions) in the medium and the light wave. It is assumed that these particles have natural frequencies of vibration, and that they obtain large amplitudes by resonance when acted on by a wave of suitable frequency. These large amplitudes involve absorption due to frictional damping. Furthermore, it may be shown either from mechanical analogies or from electrical theory that the rate of propagation of waves through a medium will be accelerated or retarded if the medium contains vibrating elements which have a free rate of vibration slightly greater or less than that of the waves. A more general dispersion formula may be derived on this basis, which for two natural frequencies, corresponding to absorption bands at λ_0 and λ_1, has the form

$$n^2 = A + \frac{B}{\lambda^2 - \lambda^2_0} + \frac{C}{\lambda^2 - \lambda^2_1}$$

This gives a discontinuity in the refractive index n where λ becomes equal to to λ_0 or to λ_1. Near these discontinuities, where the absorption is great, the

equation is not correct because it neglects frictional damping. For other regions, however, it holds quite accurately.

Selective reflection is observed at wave-lengths corresponding to the natural frequencies of the medium, if the damping is not too large. We may consider this as a re-radiation of electromagnetic waves by the charged particles, just as a tuning fork re-radiates sound waves after being excited by resonance. For these colors the reflecting power is exceptionally high, and the substance is said to show "metallic" reflection. Quartz has a short region of metallic reflection in the infra-red at 85,000 Ångström units, and sylvite at 611,000 Ångström units. Selective reflection is responsible for the colors of substances which show surface color (§722).

713. Kirchhoff's Law of Radiation.—If the fraction A of the radiation of a given wave-length incident on a body is absorbed, A is said to be its absorbing power for that color. The *emissivity* of a radiating body is the amount of energy radiated per second from each unit of surface. Kirchhoff showed by the theory of exchanges (§331) that the emissive and absorptive powers of all bodies at the same temperature for a given color are proportional when the radiation is a pure temperature effect.

714. Origin of the Fraunhofer Lines.—A general account of these lines has been given in §649.

Kirchhoff, noting that there were coincidences between many of the Fraunhofer lines and emission lines, explained them as the result of selective absorption by vapors in the sun's atmosphere of waves which these vapors may themselves emit. Stokes independently suggested that the coincidence of the yellow sodium lines with the D lines indicated that the sodium atoms must absorb waves of the same frequency as those emitted by them, the effect being similar to resonance phenomena in sound. This *reversal* of the sodium lines is easily secured by igniting a small piece of metallic sodium in a metal spoon before a slit illuminated with the electric arc, the light then passing through a prism and a lens which focuses it on a screen. Furthermore, if a large quantity of sodium vapor is present in an arc the phenomenon of *self-reversal* is shown in the spectrum. The D lines are very broad and intense, with a narrow dark line in the middle of each, due to absorption by the cooler sodium vapor in the outer portion of the arc.

715. Luminescence.—In all cases where radiation is purely a temperature effect Kirchhoff's law appears to hold. In many cases, such as those of fluorescence and phosphorescence (§§727, 728), in which the absorption of waves of certain lengths causes the

emission of waves of a different length, this is not true; nor is it generally true of luminous gases and vapors, where the luminosity appears to be due to electrical or chemical causes. In no known case do gases or vapors have absorption lines corresponding to all the emission lines. The name *luminescence* has been applied to the various kinds of radiation not directly due to high temperature and not conforming to Kirchhoff's law.

716. Solar Spectrum.—The wave-lengths of many thousands of the Fraunhofer lines were determined by Rowland. A large number were found to coincide with the emission lines of known elements, and at the present time 58 elements have definitely been identified in the solar spectrum. (See Fig. 602, which shows the coincidence of many of these absorption lines with the emission lines of iron.) The Fraunhofer lines are produced by the lower part of the solar atmosphere, known as the reversing layer. The chromosphere, or upper solar atmosphere, the prominences or flames of incandescent hydrogen and other gases rising out of it, and the corona, or nebulous outer envelope, give bright line spectra which may be seen during a total eclipse, when the brighter light from the photosphere does not mask them. The rare gas helium was known to exist in the sun before it was found on the earth, on account of the bright yellow line due to it observed in the spectrum of the prominences.

Wave-lengths of Fraunhofer Lines

A	7594–7621	O_2	Red (Band)
B	6870–6884	O_2	Red (Band)
C	6562.816	H	Red
D_1	5895.944	Na	Orange
D_2	5889.977	Na	Orange
E_2	5269.552	Fe	Green
F	4861.344	H	Blue
G	4226.742	Ca	Violet
H	3968.494	Ca^+	Violet
K	3933.684	Ca^+	Violet

The ultra-violet region of the solar spectrum does not extend beyond a wave-length of about 3000 Ångström units. Without doubt shorter waves are emitted, but they are absorbed by the earth's atmosphere, which is opaque to all very short waves. The atmosphere also exercises general and selective absorption in the visible region. Molecular oxygen gives rise to the terrestrial

bands known as the Fraunhofer lines A, α and B, and there is more or less general absorption due to these and other constituents of the earth's atmosphere.

The infra-red region of the solar spectrum has been investigated by Langley with the bolometer, and found to extend beyond a wave-length of 53,000 Ångström units. Broad absorption bands are found, some of which coincide with those due to water vapor and carbon dioxide, besides many narrow lines and bands of unknown origin. A large porportion of the solar radiation, particularly in the neighborhood of the shorter waves, is absorbed by the earth's atmosphere, and this must greatly influence climatic conditions.

717. Spectra of Planets, Stars, Comets, and Nebulæ.—The planets and the moon give spectra similar to that of the sun, since their light is merely reflected sunlight, but modified by general and selective absorption in the cases of the planets which have an atmosphere. Most stars have characteristic absorption spectra resembling that of the sun, which shows the universal distribution of many of the common elements. In addition there are frequently lines of unknown origin. Nebulæ give bright line spectra, the

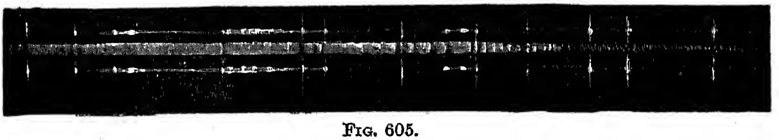


FIG. 605.

main lines being due to hydrogen, helium, and ionized oxygen and nitrogen. The spectrum of comets consists mostly of bands due to CH, C_2, CO, and CN, similar to those given by the Bunsen flame. It seems evident in the cases of nebulæ and comets that the radiation is an example of luminescence, or luminosity due to other causes than high temperature, because these bodies appear to consist of masses of highly attenuated gases, or small bodies, and it is inconceivable that their temperature can permanently remain much higher than that of the surrounding space.

718. Application of Doppler's Principle.—If a star is approaching or receding from the earth, the effect will be to shorten or lengthen each wave reaching the earth (§596). Each line will be displaced toward the violet if the star is approaching, toward the red if it is receding. By measuring such displacements on photographs of stellar spectra the velocities of stars in the line of sight may be determined with an error of less than one kilometer per second. Most of the stars which have been investigated have velocities with respect to the sun of between one and one hundred kilometers per

second. It is found that a majority of the stars on one side of the heavens have a general relative motion toward the sun, those on the opposite side away from the sun. The inference is that the solar system is itself moving through the universe in the former direction.

Fig. 605 is the spectrum of χ Draconis, with comparison spectra of hydrogen above and below, showing the Doppler effect on the hydrogen absorption lines.

EFFECTS DUE TO ABSORPTION

719. Color of Natural Objects.—The colors seen in the spectra produced by dispersion or by interference are pure. This is not the case with the colors of natural objects, which as a rule are due to selective absorption of certain colors of the incident light, the other colors being diffusely reflected in different proportions. If a colored object, such as a red rose, is placed in different parts of a spectrum, it will appear a brilliant red in the red and almost black in other parts. This shows that the greater part of all colors except red is absorbed; not all, however, for it will be noticed that in every part of the spectrum there is some reflection of the incident color. Since the resultant of the combination of all colors is white, it may thus be proved that from all colored objects some white light is reflected, in addition to the characteristic color.

720. Body Color.—In most cases it is observed that bodies having a certain color by reflected light have the same color by transmitted light. This suggests that the color diffusely reflected is due to components of the incident white light which have penetrated more or less into the medium before being scattered, the other colors being lost by absorption. The white light reflected is probably due both to reflection at the surface and to the recombination of the various colors that escape complete absorption. As a crude illustration of body color, if light falls on a piece of red glass a white image of the source will be reflected from the front surface and a red image from the rear surface.

Colors are said to be more or less *saturated* according to the proportion of white light with which they are diluted. The pure spectral colors are said to be completely saturated. The proportion of white light scattered is increased by any process which increases the reflecting surface. For example, crystals of copper sulphate will appear lighter and lighter as they are crushed into smaller fragments,

and become almost white when reduced to a fine powder. The white light reflected from the numerous surfaces then completely masks the small portion which is selectively transmitted. Similarly, transparent substances such as glass are white when in powdered form.

721. Dichromatism.—Some substances when examined by light transmitted through thick layers appear to be of different color from that observed by reflection or by transmission through a thin layer. A thin layer of chlorophyll is green by transmitted light, while a thick layer is red. This is explained by the fact that the absorptive power, or the fraction of the incident light absorbed by a layer of unit thickness, is different for the two colors. While the incident green light is more intense than the red, and remains so after transmission through a thin layer, it is more rapidly cut down by absorption, so that after passing through a thick layer the red predominates. This effect is called dichromatism.

722. Surface Color.—Some substances appear of different colors by reflected and by transmitted light. Such is the case with thin films of metal and of the solid aniline colors. Gold is always yellow by reflected light, but a sheet of gold leaf thin enough to permit transmission appears green by the transmitted light. The light reflected from these substances is complementary to that transmitted. In such cases selective action seems to take place at the surface, some colors being almost wholly reflected, while the remaining colors are transmitted through a thin film. Hence the transmitted light is complementary to that reflected. Bodies exhibiting surface color retain that color when finely powdered.

723. Colors of Sky and Clouds.—Since light can reach the eye only directly from the source or by reflection from material objects, it is evident that, since the sky is not perfectly black, it must contain matter in suspension. The blue color of the sky is due partly to selective scattering by small suspended particles of dust, water, etc., but it has been shown that the molecules of a gas scatter a considerable amount of light. It is to be expected that such small particles should reflect a larger proportion of short waves than of long ones. The term scattering is used, because it seems evident that this is not a case of ordinary reflection like that from a mirror of finite size. There is an analogy in the case of sound waves; long waves pass around obstacles without deviation from their general direction, while shorter waves may be reflected. Since the shorter waves of light are scattered, the transmitted light will consist mostly of the longer waves. This accounts for the brilliant reds, oranges, and greens often observed in the western sky at sunset. The light transmitted almost tangentially through the atmosphere has been

deprived of the shorter waves, which cause a blue sky for those more immediately under the sun. These effects are intensified by the presence of a larger number of dust particles in the lower levels of the atmosphere. After the great eruption of the volcano Krakatoa in 1883 fine volcanic dust pervaded the atmosphere of the whole earth and the sunsets were especially brilliant. For the same reason lights look red when seen through smoke or fog, or through water made slightly turbid by the addition of a small quantity of milk or shellac solution. This effect is beautifully illustrated by passing a beam of light through a jet of steam issuing from a small nozzle into a stream of air previously dried by forcing it through sulphuric acid. The size of the water drops is controlled by changing the vapor pressure in the atmosphere in which the drops are formed, lower vapor pressure promoting evaporation and thus reducing the size of the drops. The colors seen by transmitted and by scattered light are complementary, the shorter waves being scattered and the longer ones transmitted.

724. Color Sensation.—The perception of a given color by the eye does not necessarily prove that the stimulus is of the corresponding wave-length. It may be the resultant effect of several different colors. For example, if the light from the red of a spectrum and from a region intermediate between the blue and the green be superimposed the resultant sensation is white, which the eye cannot distinguish from the white due to a mixture of all the colors. A similar effect is produced by the combination of violet and yellow-green. Two colors which together give the sensation of white are said to be *complementary*. It is found, further, that spectral red and green combined excite a sensation of yellow, while green and violet produce blue. All possible colors may be produced by combining red, green, and violet. According to the theory of Thomas Young, these are to be regarded as the three primary color sensations. The cones in the retina are supposed to respond or "resonate" most actively to frequencies of vibration corresponding to these colors, and all color sensations depend on the proportions of the incident energy belonging to these frequencies.

The phenomena of color sensation may be explained by assuming that in the normal eye there are three sets of nerves, one stimulated most actively by red light, another by green, and a third by violet, but each responding also more or less to waves of other frequencies as well. To indicate these effects Koenig constructed three curves (Fig. 606) on an axis representing the length

of the normal spectrum from the Fraunhofer line K in the violet to B in the red. The ordinates at each point are considered to represent the degree of excitation of the three sets of nerves respectively by light of the frequency corresponding to that point in the spectrum. The true shape of these curves is still very uncertain. The maximum sensibility of the "red" nerves is in the orange-red, that of the "green" nerves in the green, and that of the "violet" nerves in the blue-violet. The first set of nerves is also excited more or less by all colors between H and B; the green by all colors between G and C, and the violet by all colors between K and E. The color of sodium light (D) is caused by the superposition of two sensations, red and green, proportional respectively to the ordinates $D1$ and $D2$. The color of the blue line of hydrogen (F) is due to a combination of red, green, and violet sensations proportional to the ordinates $F1$, $F2$, and $F3$. In the case of color-blind persons one or more sets of nerves are missing—usually the red. To such persons, for example, sodium light would appear green. Two colors, such as red and blue-green, are com-

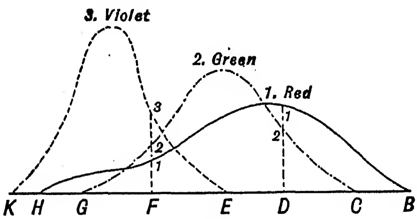


FIG. 606.

plementary, when, acting jointly, they excite all three sets of nerves in the proper proportion to produce the sensation of white—or in the same proportions that they are excited by ordinary white light.

725. Mixing Colors.—There are two distinct methods of mixing colors, one by addition, the other by subtraction. The additive method was discussed in the last paragraph. It can be illustrated by superposing beams of differently colored lights on a white screen or by rapidly rotating a disk on which are the colors to be mixed. The subtractive method is illustrated in the mixing of paints or pigments. Strips of colored gelatine laid over each other give excellent examples of this method of mixing colors. The tricolor process of printing and of color photography is based on the subtractive method of mixing colors.

If a beam of yellow light and a beam of blue light be combined on a white screen the result is white. Yellow and blue, being complementary colors, when mixed additively produce white. But if yellow and blue pigments be mixed the result is green. White light incident on the yellow pigment is diffusely reflected from various depths, but in passing through thin layers of the pigment

before and after reflection the light is mostly absorbed, with the exception of the yellow and some green which are diffusely reflected (see §§711 and 719). Similarly the blue pigment absorbs nearly all the light except blue and some green which are diffusely reflected. Hence if white light is incident on a mixture of these two pigments, green is the only color which escapes absorption by one or the other. The apparent color of pigments may vary with the kind of illumination. Blue pigments usually appear green by candle light because there is a very small proportion of blue in the light and so green predominates in the diffusely reflected light. The true colors of dry goods and carpets cannot be judged by ordinary artificial light.

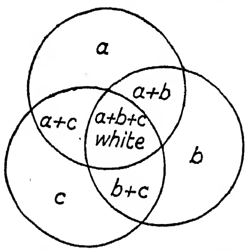


FIG. 607*a*.—Additive method.
a = red
b = green
c = blue-violet
$a + b + c$ = white

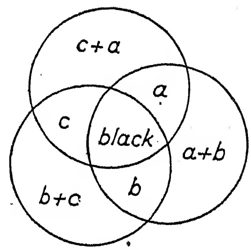


FIG. 607*b*.—Subtractive method.
$b + c$ = peacock blue
$c + a$ = crimson
$a + b$ = light yellow

The additive method of mixing colors tends to produce white while the subtractive method tends to produce black. It was pointed out in the last paragraph that there are three primary colors, red, green, and blue-violet. (The proper blue lies well toward the violet, hence some writers use violet, some blue, and some blue-violet for the third color.) There are also three primary pigments or subtractive primary colors. These are, in order, peacock blue, crimson, and light yellow. They are given in this order because each subtractive primary color is complementary to the corresponding additive primary color.

The additive and subtractive method of mixing colors may be illustrated by Figures 607*a* and 607*b*.

Fig. 607*a* illustrates the fact that by adding any two primary colors one obtains a primary pigment or subtractive primary color, and this primary pigment color is complementary to the remaining primary color. Thus the primary pigment light yellow ($a + b$) is

complementary to the primary color blue-violet. The addition of all three primary colors $(a + b + c)$ is white. Fig. 607*b* shows that by adding two primary pigments one obtains that primary color which is common to both pigments and which is complementary to the remaining primary pigment. The addition of all three primary pigments results in the absorption of all the light and hence in darkness.

726. Chemical and Molecular Effects.—Light may cause chemical combination, as when it acts on a mixture of hydrogen and chlorine, or dissociation, as when it acts on the silver salts in a photographic plate. By its action on the chlorophyll of plants, light decomposes the carbon dioxide absorbed from the atmosphere, releasing the oxygen and causing the carbon to be assimilated. It may cause molecular transformations, as when it alters amorphous to crystalline selenium, or changes the electric resistance of the latter form. It also changes white phosphorus to red. These effects are not due to the heating effect of the absorbed radiation, because an equivalent rise of temperature will not cause them, but they seem rather to depend on the transfer of the energy of light quanta (§546) to the atoms or molecules. As a rule, the shorter waves are more effective in producing such results.

Another effect due to light, especially to the ultra-violet waves. is the discharge of negative electricity from many substances (§544).

727. Fluorescence.—There are substances which when stimulated by the absorption of light waves of certain lengths will emit waves of other lengths. For example, a piece of paper moistened with sulphate of quinine solution and held in the ultra-violet portion of the solar spectrum will emit a brilliant opalescent blue light. To this phenomenon Stokes gave the name of fluorescence, because it was observed in fluorspar. He explained it as the result of the absorption of incident waves which by a modified resonance action caused a reëmission of longer waves. Similar effects are observed in coal oil, fluorescein, eosin, uranin, and other organic compounds; in uranium glass, which emits a yellowish-green light; in esculin, which emits blue light, and in chlorophyll, which emits red light; and also in a much smaller degree in iodine, wood, paper, and many other substances.

728. Phosphorescence.—There are many substances, of which calcium, strontium, zinc, and barium sulphides are familiar examples, that after exposure to light show effects which are similar to fluorescence, but which continue visible long after the exciting radiation ceases to act. This is called phosphorescence. The only definite distinction between fluorescence and phosphor-

escence is that the latter persists after the excitation has ceased, while the former does not. Many substances which phosphoresce very feebly at ordinary temperatures may be made to glow brilliantly at the temperature of liquid air. As examples, gelatin, horn, egg shells, and paper may be mentioned.

Some metallic vapors, such as those of the sodium and calcium group, fluoresce brilliantly under the action of light or cathode rays. The light shows the characteristic spectral lines and bands of the metal. Certain organic vapors, such as anthracene, fluoresce when light falls on them. Nitrogen, oxygen, and some other gases will under certain conditions phosphoresce brightly for several seconds after an electric discharge has passed through them in a vacuum tube.

RADIATION AND ATOMIC STRUCTURE

729. Origin of Line Spectra.—The explanation of the sharp lines observed in spectra (§707) presented insuperable difficulties to the older theory that light waves are due to periodic motions of the electrons in atoms. Thus if we accept the nuclear model of the atom (§540) an electron which radiates light would undergo a continuous change not only in its energy, but also in the frequency of its motion. The frequencies of light could not be confined to particular values corresponding to sharp spectrum lines, but would be spread out to give a continuous spectrum.

The first success in overcoming this difficulty was achieved by Niels Bohr, who extended the ideas put forward by Planck in deriving the law of black body radiation (§540). Planck had found it necessary to assume that the energy is radiated in definite and discrete amounts called quanta. The laws of the photoelectric effect (§544) gave independent evidence for the existence of quanta. Bohr went further in assuming that the atom itself can exist only in certain states of definite energy, and that it was impossible for it to have energy corresponding to any amount between two allowed states. In the hydrogen atom, which consists of a single electron and the nucleus, these states were pictured as certain particular orbits in which the electron could revolve around the nucleus. In order to obtain a stable atom, it was necessary to assume that no radiation is emitted when the atom is in a given state. This was in direct contradiction to the fact that the centripetal acceleration of the charged electron in its orbit should, accord-

ing to the classical laws of electricity, cause it to radiate energy. According to Bohr, light is generated by an atom only when a transfer occurs from a state of energy E_1 to one of lower energy E_2. The energy of the light constitutes a quantum, equal to the loss of energy by the atom. Since the energy of a quantum is proportional to the frequency ν of the light (§546), we obtain the relation

$$E_1 - E_2 = h\nu$$

known as *Bohr's frequency condition.* The constant of proportionality h is Planck's constant, with the value 6.55×10^{-27} when the energy is in ergs.

Bohr's theory was remarkably successful in accounting with great accuracy for the series of spectral lines observed from hydrogen. The permitted orbits were selected by a simple rule which had little justification except that when the corresponding energies were substituted in the frequency condition the correct results were obtained. Further confirmation was obtained from the spectrum of ionized helium, which also has one electron outside the nucleus. By measuring the wave-lengths of the lines of hydrogen and ionized helium, it was possible to obtain a very accurate value of the ratio of charge to mass of an electron.

The extension of the orbital model to atoms with more than one electron has been successful in a qualitative way, but has proved incapable of yielding quantitative values for the discrete energy states. The problem has been attacked in recent years by a new theory based on the wave character of the electron (§758). This method gives the correct results in all cases where the problem can be solved, and for hydrogen it yields energy states identical with those of the Bohr theory. An important feature of the wave method is the limitation that it sets upon any attempt to specify an exact orbit for the electron. The Bohr frequency condition, and the idea of stationary energy states, remain as an important part of the new theory. In fact, the selection of the allowed states is made in a much more reasonable way than that originally used by Bohr.

730. Atomic Energy States.—From the line spectrum of an element it is possible to deduce the atomic energy states by the application of the frequency condition. Having various values of ν, calculated from the observed wave-lengths of the lines, one may find values of the energies E_1, E_2, E_3, etc., such that the frequency condi-

tion is satisfied for all lines. This has been done for a great many of the elements. Fig. 608 shows the states of hydrogen and of sodium, which are among the simplest of those obtained. Each horizontal line represents a possible energy state of the atom, and the arrows show the transitions between states which give rise to some of the strongest spectral lines. Thus when the hydrogen atom changes from state 3 to state 2 it emits H_α, a strong line in the red part of the spectrum. Other jumps to state 2 give the remaining lines of the so-called Balmer series. The frequencies of the lines are proportional to the lengths of the arrows, since these represent the

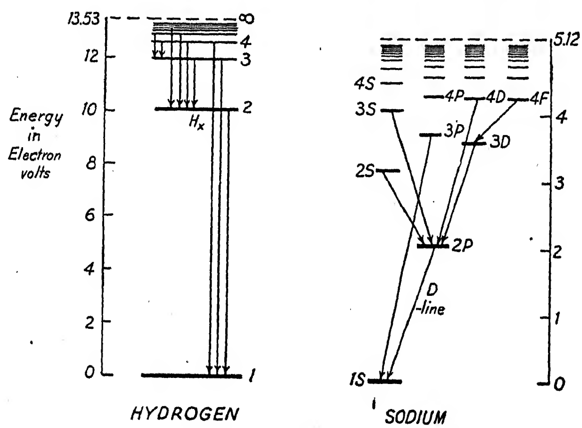


Fig. 608.

differences $E_1 - E_2$. The states for sodium are more numerous, there being several sets between which transitions may occur. All the possible transitions do not give observed lines however, and it will be noticed that arrows have been drawn only between adjacent sets of states. The strong yellow sodium line is the result of the transition between states $2P$ and $1S$.

In a gas at room temperature all of the atoms will ordinarily be in the lowest or *normal* energy state, *i.e.* state 1 for hydrogen or state $1S$ for sodium. To be able to emit light, they must first be raised to higher states by the addition of energy. There are two common ways in which this can occur. The atoms may be struck by electrons of sufficient kinetic energy, or they may absorb the energy of incident radiation.

It is found that if electrons of controlled velocity, produced by a filament and grid method, traverse sodium vapor, no radiation is

obtained from the vapor unless the electrons have fallen through a potential of at least 2.1 volts. As shown in Fig. 608, the first state to which a sodium atom can be raised is $2P$. The energy of this state equals that of an electron which has fallen through 2.1 volts, an energy which is called 2.1 electron-volts. Electrons of energy 5.12 electron-volts or more will completely remove the outer electron from a sodium atom, and cause ionization. Thus 5.12 volts is called the *ionization potential,* and no discrete states of the neutral atom exist with energies greater than this.

Absorption of radiation by the atom is the reverse of emission. If white light is passed through sodium vapor, the atoms are raised from $1S$ to the various P states by absorbing light of the appropriate frequencies. They subsequently return to the normal state, re-emitting either the frequency absorbed, or some lower frequency. Thus an atom in $3P$ could either return directly to $1S$, emitting a single line, or by way of $2S$ and $2P$, emitting three lines of lower frequency. The latter process is called fluorescence (§727).

731. X-ray Spectra.—Whereas the energy states involved in the emission of light are due to different configurations of the outermost electrons in the atom, those giving rise to X-ray spectra (§542) involve the removal of an electron from one of the stable inner shells of electrons. These shells are much closer to the nucleus and hence the states have much greater energies. For the same reason, they vary in a regular fashion with the charge on the nucleus, or the atomic number (§540). It was by a study of the line spectra of X-rays from different elements that Moseley first established the correct relative atomic numbers.

732. Molecular Spectra.—The spectra of molecules were described in §708 under the heading Band Spectra. Molecules possess discrete energy states just as do atoms. Different arrangements of the electrons give rise to states of electronic energy bearing considerable resemblance to those of atoms. But in molecules there are additional states due to the vibration of the system of atoms and to the rotation of the molecule as a whole. Changes of the vibrational and rotational energies cause the characteristic appearance of a band spectrum. A group of bands called a *band system* arises from a given change in the electronic configuration. Each individual band results from a given change in the vibrational energy of the molecule, while each of the fine lines of which the band is composed is characterized by a particular change in rotational energy.

The states of rotational energy are very closely spaced, but can be evaluated by measuring the individual lines in the bands. Their spacing depends on the rotational inertia of the molecule, and hence it is possible to determine the distances between the nuclei of the atoms of which the molecule is composed. This is now the most accurate method for obtaining molecular dimensions. The vibrational states, which may be found from measurements of the wavelengths of the various bands in a band system, depend on the frequency of vibration of the nuclei with respect to each other, and on its variation with amplitude. A study of these allows us to find the forces acting between nuclei, and the work necessary to completely separate the atoms. The latter quantity is called the heat of dissociation of the molecule. In the case of very stable molecules, such as CO, O_2 and NO, which are dissociated only at extremely high temperatures, the spectroscopic determination of the heat of dissociation constitutes the only feasible method. Another recent application of band spectra is the discovery of isotopes (§539) of oxygen, carbon and nitrogen. While the line spectra of isotopes are practically identical, because of the identical nuclear charge, molecules containing different isotopes show different band spectra due to the effect of mass on the rotational inertia and vibration frequency. Thus band spectra furnish a valuable method for the study of isotopes.

DOUBLE REFRACTION AND POLARIZATION

733. Double Refraction.—Some crystals, such as those of rock salt and fluorite, resemble isotropic solids, such as glass, in the respect that their physical properties are alike in all directions. In general, however, this is not the case; such properties as elasticity and heat conduction, as well as optical properties, differ in different directions in the crystal. In such crystals as quartz and calcite there is an *axis of symmetry*, the crystallographic axis, and the physical properties are the same in all directions in any equatorial plane, but different from those in the direction of the axis. Iceland spar, or calcite, is a rhombohedral crystal, each face being a parallelogram with two acute angles of 78° 5′ and two obtuse angles of 101° 55′. Two solid angles of the crystal are formed by the junction of the obtuse angles of three faces. Any line equally inclined to the faces of one of these solid angles is a crystallographic axis. An object seen through Iceland spar appears double, unless

viewed in the direction of the axis. No such effect is observed in the case of isometric crystals. This phenomenon is called *double refraction.* When the waves travel in the crystal in the direction of the crystallographic axis there is no double refraction; hence *any line in the crystal parallel to the crystallographic axis* is called an *optic axis.*

If a ray i of ordinary light is incident normally on any face of a doubly-refracting crystal one ray o is transmitted without deviation; and if the incidence is oblique this ray is deviated with an index of refraction which is independent of the angle of incidence. The other ray e is deviated in all cases, unless it travels along an optic axis, and the index of refraction varies with the angle of incidence.

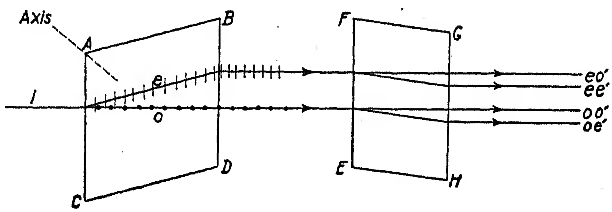


FIG. 609.

dence. The first is called the *ordinary,* the second the *extraordinary ray.* If the incident light is normal (Fig. 609), and if the crystal be rotated, keeping the surface AC parallel to itself, the ordinary image will remain at rest, while the extraordinary image rotates about it in such a way that the line joining the two images lies in a *principal section,* a plane including the normal to the surface and an optic axis. If the ordinary and extraordinary rays o and e pass normally through a second crystal each ray generally divides in two, the rays oo' and oe' and the rays eo' and ee' (Fig. 608), the line joining each pair lying in a principal section of the second crystal. This gives rise to four images of the source, which are of equal intensity when the principal sections of the two crystals are at an angle of 45° with each other. If this angle be changed one pair of images will increase in intensity and the other diminish. When the principal sections are parallel only the rays oo' and ee' emerge; when they are at right angles, only the rays oe' and eo'. From such experiments Huygens recognized the fact that light which has passed through Iceland spar, quartz, and other doubly refracting crystals does not possess properties which are alike in all azimuths

around the direction of propagation. Newton, in order to explain this, supposed the light corpuscles to be endowed with polarity of some sort—hence the name polarized light.

734. Direction of Vibration.—Fresnel explained the phenomenon of double refraction as a result of the transverse vibration of light waves. If the vibrations were longitudinal, it is impossible to conceive how they could be affected by rotation of the crystal in a plane at right angles to the direction of propagation. Transverse vibrations in a cord may be said to be polarized. Such vibrations would be freely transmitted through a slot parallel to the direction of vibration, but not through one at right angles to this direction. Longitudinal vibrations in a cord could be freely transmitted through a slot, regardless of its position. Fresnel assumed that in ordinary white light successive waves reaching a given point of space vibrate in different planes at random, so that, although each individual wave is vibrating transversely in a definite plane, and is, therefore, polarized, this direction changes so rapidly that the eye cannot take account of it and no polarization effects are observed. In passing through a doubly-refracting crystal vibrations in one direction travel with a different velocity from those in another direction, on account of the difference of the physical properties of the crystal in these directions, hence double refraction results. The displacement in each wave is in general resolved into two components, unless the light is travelling parallel to the axis. In that case it is unmodified, as the velocity of propagation is independent of the azimuth. In the ordinary ray, which travels in all directions with the same velocity, the vibrations must be at right angles to the optic axis. So long as this is the case the displacements will take place under the same conditions in every azimuth and the velocity be unchanged. The vibrations are perpendicular also to the ray and hence are *perpendicular* to a plane defined by the *ray* and the *optic axis*. Such a plane is known as a *principal plane*. In the case of the extraordinary ray the relations are more complex (§742), but all the experimental facts may be accounted for by assuming that the vibrations in the extraordinary ray are *parallel* to *its principal plane*, i.e., a plane formed by the extraordinary ray and the optic axis. In the case of normal incidence (Fig. 609) the principal planes of the ordinary and extraordinary rays coincide with each other and with the principal section. In this figure the optic axis for the first crystal is assumed

to be in the plane of the paper, so that this is the common principal plane and principal section. The direction of vibration of the rays *e* and *o* is indicated, that for *o* being perpendicular to the plane of the paper.

735. The Wave Surfaces.—From the experiments described above it may be seen that a wave of ordinary light on entering a doubly-refracting crystal is divided into two waves, one of which, *o*, has the same velocity in all directions in the crystal. The other wave, *e*, has a velocity which varies in different directions, and is the same as that of the ordinary wave only when both travel in the direction of the optic axis. Huygens showed that these facts are consistent with the existence of a double wave surface in the crystal, a sphere and an ellipsoid of revolution, which are tangent to each other at the two points where they intersect an optic axis. In one class of crystals, like Iceland spar, the sphere is inside the ellipsoid, and the ordinary wave is the more refracted (Fig. 610*a*). In another class, represented by potassium sulphate or quartz, the sphere encloses the ellipsoid, and the ordinary ray is less refracted (Fig. 610*b*). The first are called negative and the second positive crystals.

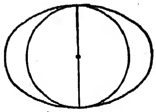


FIG. 610*a*.

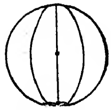


FIG. 610*b*.

In crystals in which the physical properties are different along three axes at right angles to each other, such as sugar and topaz, which likewise show double refraction, there are two axes of no double refraction; hence such crystals are said to be *biaxial*, as contrasted with the class described above, which are said to be *uniaxial*. Both rays in biaxial crystals are extraordinary, that is to say, do not conform to the ordinary laws of refraction.

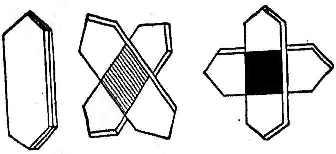


FIG. 611.

736. Double Refraction by Tourmaline.—Tourmaline is a semi-transparent hexagonal crystal. If light falls on a crystal, part is transmitted. If this falls on a second plate with its optic axis parallel to that of the first (Fig. 611), some of the light gets through; but if the second crystal is rotated about the line joining the two, less light gets through, and when its axis is at right angles to that of the first none is transmitted. Evidently the waves have had their

mode of vibration so changed by passage through the first plate that they cannot pass through the second unless the principal planes of the two are parallel. If the light first passes through Iceland spar it is found that the extraordinary ray alone will pass through tourmaline if the principal planes of the Iceland spar and tourmaline are parallel, the ordinary ray alone if they are at right angles. It follows that light is doubly refracted by tourmaline, but that the ordinary ray is totally absorbed. Recently it has been found possible to produce large sheets composed of oriented microscopic crystals embedded in nitrocellulose which produce the same effect as a single crystal of tourmaline, but with less general absorption.

As a remarkable example of Kirchhoff's law (§713), it may be mentioned that if tourmaline is raised to a high temperature it emits polarized radiation. If this falls on a second crystal parallel to the first it is absorbed, showing that it corresponds to the ordinary ray. The mode of vibration which is absorbed corresponds to that which is emitted.

737. Polarization by Reflection.—About 1808 Malus discovered that light reflected from glass at a definite angle, known as the polarizing angle, acquires properties similar to that of light transmitted through tourmaline or Iceland spar. When light is thus polarized by reflection from a mirror A (Fig. 612a) about fifteen per cent is reflected from another mirror B if the angle of incidence is the same as for A and if the two planes of incidence coincide.

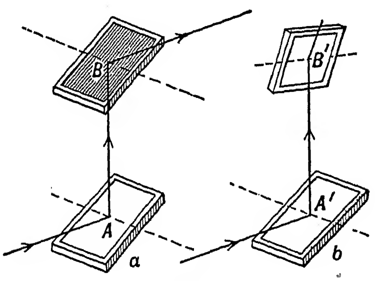


FIG. 612.

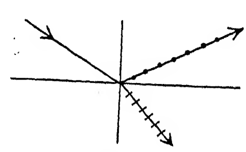


FIG. 613.

If the planes of incidence are at right angles (Fig. 612b) practically none is reflected. If the light reflected from a glass plate is examined through a crystal of Iceland spar, the ordinary ray alone is transmitted when the plane of reflection coincides with a principal plane, the extraordinary ray alone when the two are at right angles. In intermediate positions portions of both rays are transmitted. Similarly, light reflected from glass is not transmitted through tourmaline if the plane of reflection is parallel to the optic axis of the crystal.

The simplest explanation of these effects seems to be that when light strikes a reflecting surface there is a partial resolution into components respectively in and at right angles to the plane of incidence. The vibrations parallel to the surface are most freely reflected, while the others strike down into the surface and are transmitted or absorbed (Fig. 613). If polarized light is incident at the polarizing angle (§740) on a piece of glass a certain proportion will be reflected when its vibrations are parallel to the surface; if the vibrations are in the plane of incidence it will be almost wholly refracted. In general both components are reflected and refracted, but the reflected light contains a larger proportion of waves vibrating perpendicularly to the plane of incidence, and the refracted light a larger proportion of the waves vibrating parallel to that plane.

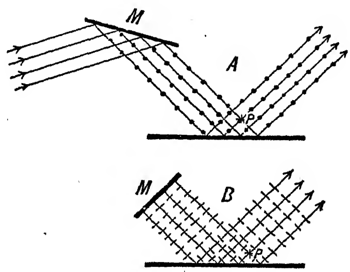


FIG. 614.

It is clear that no interference effects can be produced between two vibrations in planes at right angles to each other. This fact enabled Wiener to determine the direction of vibration in light polarized by reflection. A beam polarized by a glass plate M fell at an angle of 45° on a thin transparent photographic film above a metal reflecting surface. He found that stationary waves (§675) were produced when the plane of incidence on the film coincided with the plane of reflection from the mirror (Fig. 614A), but this was not the case if the glass plate M was rotated about the incident light, keeping the angle of incidence constant, until the two planes were at right angles to each other (Fig. 614B). From the figure it appears that in the first case the vibrations must have been parallel to the film, and therefore to the mirror, in order that the incident and reflected rays should be in a condition to interfere at P, while in the second case the vibrations must have been in the plane of incidence on the film, and therefore parallel to the mirror, in order that the vibrations should meet at P at right angles to each other. This demonstrates that the vibrations in light polarized by reflection are parallel to the glass surface.

738. Plane Polarized and Ordinary Light.—The experimental evidence seems to indicate that light waves are excited by displacements of the particles of material sources, these particles being probably ions or electrons within the molecules (§416); that these particles in general move in different planes and directions, and that the displacements of a given particle may constantly change in direction; that each particle sends out into the surrounding medium a series of waves vibrating in the plane of motion of the particle,

so that ordinary white light consists of a mixture of waves of many lengths, the resultant vibrations being in a plane at right angles to the direction of propagation, successive trains of waves having different planes of vibration; and that by refraction or reflection we may sift out component vibrations in a given plane, and produce what is called polarized light. *When all the vibrations are in parallel planes the light is said to be plane polarized.* If such light is mixed with ordinary light, it is said to be *partially polarized.* If phase differences are introduced between two vibrations at right angles, the resultant displacement may be elliptical or circular (§243). This gives rise to *elliptically* or *circularly* polarized light.

739. Plane of Polarization.—Before the direction of vibration in polarized light was known it became customary to speak of the "plane of polarization" of a polarized beam, rather than of the direction of vibration, and this plane was so defined that it coincides with the plane of incidence when the light is polarized by reflection. It follows that **the vibrations in a polarized beam are at right angles to the plane of polarization.**

740. Brewster's Law.—The light reflected from a surface is not in general completely polarized, that is, all its vibrations are not strictly in one plane. It is found, however, that for each reflecting substance there is a certain angle of incidence for which the polarization is a maximum. This is called the *polarizing angle.* It was found by Fresnel that complete polarization is given only by substances having an index of refraction equal to about 1.46. Brewster found that the polarizing angle is such that the reflected and the refracted rays are at right angles to each other. Since $n = (\sin i)/(\sin r)$ and since, when $i = p$, the polarizing angle, $p + r = 90°$,

$$n = \frac{\sin p}{\cos p} = \tan p$$

From this the polarizing angle p can be found. This is Brewster's law.

When the angle of incidence is slightly different from that defined by this equation, and even for that angle when the index of refraction differs appreciably from 1.46, a small part of the component vibrating in the plane of incidence is reflected, with a phase different from that of the other component, resulting in elliptically polarized light (§750).

741. Pile of Plates.—Since only a small fraction of the incident light is reflected from a transparent substance, even when the reflected light is completely polarized, that which is refracted will be only partially polarized; that is to say, along with light vibrating in the plane of incidence a considerable proportion of that vibrat-

ing at right angles to this plane will be transmitted. If it is subject to a second reflection the proportion of polarized light is increased. After passing through eight or ten plates the transmitted light is almost completely polarized. If a pile P of thin glass plates is built up as shown in Fig. 615, the beam R, the result of successive reflections, and the beam T, which is transmitted, are almost completely polarized in planes at right angles to each other and are of about equal intensity. This is one of the simplest methods of securing polarized light.

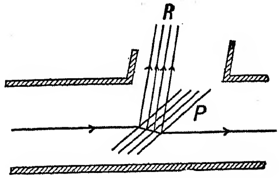


FIG. 615.

742. Wave Front Construction.—Refraction by a plane surface of doubly-refracting material may be investigated by applying Huygens' principle (§627). Here, however, the spherical wavelets which were used for isotropic media must be replaced by the double wave surfaces described in §735. In Fig. 616 Let MA and NA' be two rays corresponding to a plane wave incident normally on a crystal of Iceland spar in which the optic axis has the direction indicated. By Huygens' construction the new wave fronts, after any interval of time, are found by drawing tangent planes to the

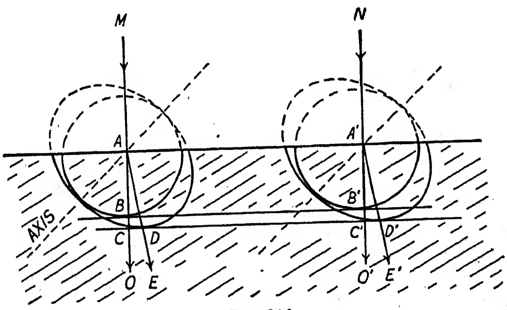


FIG. 616.

series of wavelets spreading out from each point of the surface. If A and A' are two such points, those portions of the circles and ellipses inside the crystal represent the corresponding double wavelets. Evidently the ordinary wave front BB' is a plane parallel to the original wave front, and AB represents its velocity if the wave front is assumed drawn after unit time has elapsed. Similarly DD' is the extraordinary wave front. It is parallel to BB' and tangent to

the two ellipses. In the extraordinary wave the energy moves along an extraordinary ray such as AD, with the ray velocity AD, while the normal velocity of the wave front has the smaller value AC. In this case, then, the direction of propagation of the energy (the ray) is *not perpendicular to the wave front.* This is the condition of affairs shown in the crystal $ABCD$ of Fig. 609.

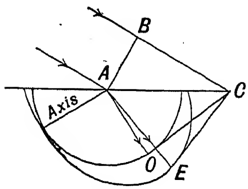


FIG. 617.

The direction of the vibration in the ordinary ray is perpendicular to its principal plane (§734), which in Fig. 616 is the plane of the paper. The extraordinary vibration is in the principal plane for the extraordinary ray and parallel to DD' (transverse wave) as shown in the figure. When the wave fronts are parallel, the principal planes for the ordinary and extraordinary rays coincide, and are identical with the principal section.

The construction of the refracted wave front for any angle of incidence may also be accomplished by using Huygens' principle as in §643. For example, Fig. 617 illustrates a case where the optic axis lies in the plane of incidence. The plane wave AB starts two wavelets at A the disturbances from which have reached O and E respectively while the wave has traveled from B to C in air. The tangent planes CO and CE are the two wave fronts. When the axis is parallel to the surface, but at right angles to the plane of incidence, as in Fig. 618, the extraordinary wave also has a circular section. Only in this plane of incidence is the ratio $\dfrac{\sin i}{\sin r}$ constant for the extraordinary ray, and this ratio is called n_e, the extraordinary index of refraction. The value of the ratio V/V_e varies with direction in every other plane of incidence, and hence cannot properly be called an index of refraction. In Fig. 618 the principal plane of the ordinary ray is perpendicular to the paper and cuts it in AO. The ordinary vibration is perpendicular to this principal plane and hence in the plane of the paper, perpendicular to AO. The principal plane of the extraordinary ray is also perpendicular to the paper, but cuts it in AE. The extraordinary vibration is in this plane, and in this case is parallel to the optic axis and therefore perpendicular to AE. Hence the two vibrations are perpendicular, although their principal planes do not coincide. In the most general case where the optic axis is oblique to the surface and not in the plane of incidence, the

extraordinary ray is also not in this plane. The two principal planes and the principal section are then different from each other and from the plane of incidence.

743. Uniaxial Prisms.—When light is incident on a doubly-refracting prism with its axis parallel to the refracting edge (Fig. 619), the ordinary and the extraordinary rays will be separated, and the angular divergence will persist after emergence. Two spectra will be formed, with light polarized in perpendicular planes.

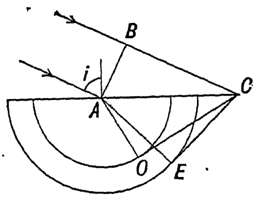


FIG. 618.

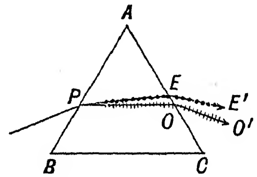


FIG. 619.

The ordinary spectrum will be less deviated than the extraordinary by a quartz prism and more by a calcite prism. When the optic axis is parallel to the refracting edge of the prism the two indices of refraction may be determined from the relations

$$n_o = \frac{\sin \frac{1}{2}(A + D_o)}{\sin \frac{1}{2}A} \text{ and } n_e = \frac{\sin \frac{1}{2}(A + D_e)}{\sin \frac{1}{2}A}$$

Some values of the indices of refraction for sodium light are given below:

	n_o	n_e
Positive Crystals:		
Quartz	1.5442	1.5533
Ice	1.3091	1.3104
Negative Crystals:		
Calcite (Iceland spar)	1.6584	1.4864
Beryll	1.5740	1.5674
Sodium nitrate	1.5874	1.5361

The difference between n_o and n_e is greater in the case of Iceland spar than in any other ordinary crystal.

744. Polarizing Prisms.—The two polarized rays produced by a doubly refracting crystal are not sufficiently separated to be conveniently used when a single beam is desired. The separation may be increased by using an ordinary triangular prism, but this introduces dispersion, so that other devices must be employed. The most common is the rhombohedral prism invented by Nicol, of Edinburgh, in 1828. In the principal section of a crystal of calcite (Fig. 620) the angles at B and D are 71°. The two end faces AB and CD are cut down to $A'B$ and $C'D$, so that these angles are

reduced to 68°. The crystal is then sliced along $A'C'$ in a plane perpendicular to the ends and to the principal section (plane of the paper). The two surfaces are polished and cemented together with Canada balsam, which has an index of refraction (1.55) less than that of the calcite for the ordinary and greater than that for the extraordinary ray (§743). If a ray of light r is incident in a direction parallel to the edge AD the ordinary ray will be totally reflected from the Canada balsam, while the greater portion of the extraordinary ray will be transmitted. The reduction of the angles at B and D is for the purpose of securing the proper angle of incidence on the balsam to produce this effect. The wave front construction in this case is like Fig. 617, and the principal section coincides with both principal planes. Hence the vibrations in the emergent extraordinary ray are parallel to the plane of the paper, as shown, and are parallel to the short diagonal of the diamond-shaped end of the Nicol.

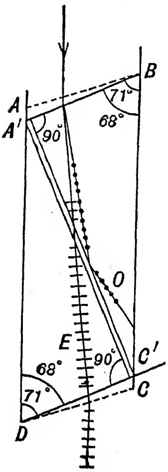


Fig. 620.

The Foucault prism resembles that of Nicol, but the total reflection is from an air film. This allows the prism to be made shorter, but there is a greater loss of light by reflection and a smaller field of view.

745. The Polariscope is an instrument for the study of the optical properties of substances with respect to polarized light. It consists of two Nicol prisms or piles of plates, one called the *polarizer*, to produce the polarized light, the other the *analyzer*, which may be set with its principal section at any desired angle with that of the polarizer, to test the incident light with respect to the nature and direction of its polarization. If any doubly refracting substance is placed between the two its effects on the polarized light transmitted through it may be studied by the analyzer.

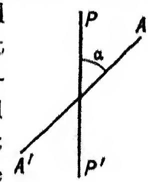


Fig. 621.

746. Resolution and Composition of Vibrations.—If the polarizer and analyzer are set with their principal sections parallel, light which has traversed the first will pass through the second without sensible loss. If their principal sections are at right angles to each other, or "set for extinction," no light will be transmitted through the analyzer. If the angle between the principal sections

is α (Fig. 621), and if a is the amplitude of the light transmitted by the first Nicol, the amplitude of that transmitted through the second is $a \cos \alpha$, and its intensity is proportional to $a^2 \cos^2 \alpha$. The intensity of the totally reflected ordinary ray is $a^2 \sin^2 \alpha$. The sum of the two intensities is $a^2(\cos^2 \alpha + \sin^2 \alpha) = a^2$, which is equal to the intensity of the light incident on the analyzer. This simple law of resolution of vibrations into components by double refraction, giving determinate control of the intensity through a wide range, is made use of in several forms of photometer.

If the two Nicols are replaced by two crystals of calcite with their principal sections at an angle of α with each other, as in Huygens' experiment (§733), an ordinary ray o and an extraordinary ray e of the same amplitude a are produced by the resolution of the vibrations along two directions in the first crystal. At incidence on the second crystal, the ordinary ray will be resolved into the components oo' and oe' of amplitudes $a \sin \alpha$ and $a \cos \alpha$ and the extraordinary ray into the components eo' and ee', of amplitudes $a \sin \alpha$ and $a \cos \alpha$. There will be, therefore, in general four rays, as found by Huygens, which will be of equal intensity when $\alpha = 45°$. When the principal sections are at right angles, the incident ordinary ray goes through the second crystal as an extraordinary ray and the extraordinary as an ordinary ray, and there are only two images.

If the second crystal is replaced by a Nicol prism, with its principal section at an arbitrary angle to that of the first crystal, only the components oe' and ee' emerge, their vibrations being in the same plane, that of the principal section of the analyzer. If the two rays are superimposed on emergence, the intensity will depend not only on the amplitudes of the two components, but on the phase differences which have been introduced owing to the difference in velocity in the crystal of the two rays from which these components are derived; in other words, there may be interference provided the light falling on the first crystal is plane polarized (see next section.)

747. Interference of Parallel Polarized Light.—If parallel plane polarized white light passes through a doubly refracting crystal of uniform thickness t and then through an analyzer, uniform colored effects are produced over the entire field, since some colors are reënforced and some weakened by interference. There is no real loss or gain for any color, for, as shown in §746, whatever energy is lost in the extraordinary ray is gained by the ordinary, and conversely,

is α (Fig. 621), and if a is the amplitude of the light transmitted by the first Nicol, the amplitude of that transmitted through the second is $a \cos \alpha$, and its intensity is proportional to $a^2 \cos^2 \alpha$. The intensity of the totally reflected ordinary ray is $a^2 \sin^2 \alpha$. The sum of the two intensities is $a^2(\cos^2 \alpha + \sin^2 \alpha) = a^2$, which is equal to the intensity of the light incident on the analyzer. This simple law of resolution of vibrations into components by double refraction, giving determinate control of the intensity through a wide range, is made use of in several forms of photometer.

If the two Nicols are replaced by two crystals of calcite with their principal sections at an angle of α with each other, as in Huygens' experiment (§733), an ordinary ray o and an extraordinary ray e of the same amplitude a are produced by the resolution of the vibrations along two directions in the first crystal. At incidence on the second crystal, the ordinary ray will be resolved into the components oo' and oe' of amplitudes $a \sin \alpha$ and $a \cos \alpha$ and the extraordinary ray into the components eo' and ee', of amplitudes $a \sin \alpha$ and $a \cos \alpha$. There will be, therefore, in general four rays, as found by Huygens, which will be of equal intensity when $\alpha = 45°$. When the principal sections are at right angles, the incident ordinary ray goes through the second crystal as an extraordinary ray and the extraordinary as an ordinary ray, and there are only two images.

If the second crystal is replaced by a Nicol prism, with its principal section at an arbitrary angle to that of the first crystal, only the components oe' and ee' emerge, their vibrations being in the same plane, that of the principal section of the analyzer. If the two rays are superimposed on emergence, the intensity will depend not only on the amplitudes of the two components, but on the phase differences which have been introduced owing to the difference in velocity in the crystal of the two rays from which these components are derived; in other words, there may be interference provided the light falling on the first crystal is plane polarized (see next section.)

747. Interference of Parallel Polarized Light.—If parallel plane polarized white light passes through a doubly refracting crystal of uniform thickness t and then through an analyzer, uniform colored effects are produced over the entire field, since some colors are reënforced and some weakened by interference. There is no real loss or gain for any color, for, as shown in §746, whatever energy is lost in the extraordinary ray is gained by the ordinary, and conversely,

so that the ordinary light which is internally reflected in the prism is complementary to that passing through. When the principal sections of the crystal and the analyzer are either parallel or at right angles to each other no modification of the light is produced, the ordinary or the extraordinary ray alone getting through, so that there can be no interference. In all other positions of the analyzer there are varying proportions of white and colored light transmitted, the color effects being most pronounced when the principal sections are at an angle of 45° with each other.

The original beam of light falling on the crystal must be plane polarized. If ordinary light is used the succession of waves vibrating in different planes when resolved in the crystal will give rise to all possible distributions of amplitudes, so that all colors will be equally affected and the resultant effect will be white light.

748. Double Refraction Due to Strain.—If a plate of glass or other isotropic substance is placed between a polarizer and an analyzer set for extinction no effect is produced. If the substance is then compressed or stretched some light will pass and interference effects similar to those described above will be produced. This shows that an istotropic substance becomes doubly refracting when subjected to non-uniform strain. This method offers a very delicate test of deviations from isotropy. Some liquids show the same characteristics in cases where the viscosity is so great or the stress so suddenly applied that a uniform hydrostatic pressure has not had time to become established throughout the substance. Imperfectly annealed glass exhibits double refraction. As shown by Tyndall, a bar of glass set in longitudinal vibration restores the light through the crossed Nicols, and a rotating mirror shows that the effect is set up periodically as the compression waves pass across the field.

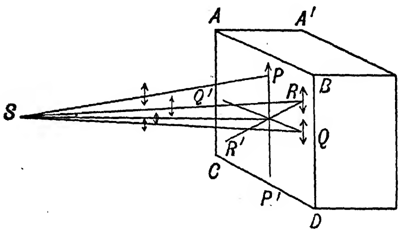


Fig. 622.

Kerr found that a block of glass in a strong electrostatic field becomes doubly refracting like a uniaxial crystal with its axis parallel to the field.

749. Interference of Convergent or Divergent Light.—If a divergent or convergent pencil of polarized light falls on a doubly refracting crystal different

portions of the pencil will traverse the crystal at different angles, and therefore with different optical paths; hence the interference effects will not be uniform over the whole field. In general the effects are quite complex and cannot be discussed here, but the simple case of a uniaxial crystal cut perpendicularly to the optic axis may be considered as an illustration. Consider such a pencil diverging from S and falling normally on the face $ABCD$ of a doubly refracting crystal with its axis parallel to AA' (Fig 622). The vibrations of the incident light may be supposed to be in a vertical plane, as indicated by the arrows. At P and Q the incident vibrations are respectively parallel and perpendicular to the principal planes PP' and QQ' and travel through without change If an analyzer is placed beyond the crystal and set for transmission or extinction of light transmitted by the polarizer, there will be a light cross or a dark cross on a screen beyond it corresponding to the crossed lines PP' and QQ'. The light incident at such a point as R, however, will be vibrating at an angle with the principle plane RR', and will be resolved into two components. A relative difference of phase between them will exist at emergence, and interference effects will take place when they are re-resolved into the same plane by the analyzer The same difference of path will exist for all rays incident at the same angle, that is, at all points equidistant from the normal from S to $ABCD$, hence colored rings similar to Newton's rings in appearance will be projected on a screen beyond the analyzer The "rings and brushes" due to a calcite plate are shown in Fig. 623. The brushes are dark, showing that the Nicols are crossed.

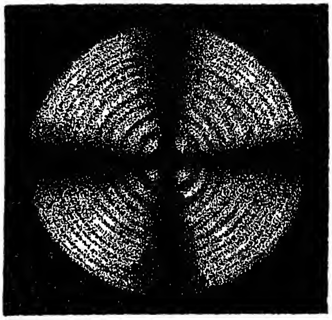


Fig. 623.

The interference effects due to crystals cut in other ways or to biaxial crystals are analogous to those described above, but more complex.

750. Circular and Elliptical Polarization.—Consider the state of the light originally plane polarized as it emerges from a doubly refracting crystal before it reaches the analyzer. The ordinary and extraordinary rays start from the first surface in the same phase, but, as their velocities are different, one set of waves will fall behind the other At different points within the crystal there will be two vibrations at right angles to each other and with phase differences depending upon the thickness of the medium traversed. The optical difference of path d at a distance t from the first surface is $[(V/V_e) - (V/V_o)]t$. At points where this difference is $N\lambda$ the light is plane polarized in a direction intermediate between the planes of vibration of the two components, the slope depending on their relative amplitudes, and being 45° if these are equal If the difference of path is $(N + \frac{1}{2})\lambda$ the light will likewise be plane polarized, but with a reversed direction of slope If the difference of path is any odd multiple of a quarter of a wave-length the disturbance will be elliptical, or circular if the amplitudes are equal. For intermediate differences of path the disturbance will be elliptical, the axes of the ellipse being oblique with

respect to the axis of the crystal. The successive forms for different path differences are shown in Fig. 624, the vibration at $d = 0$ representing the original incident light. On emergence from the crystal the disturbance will preserve the final form, and will be plane, elliptically, or circularly polarized according to the thickness of the crystal. If the waves are circularly polarized the disturbance travels through space like a point on a rotating screw. The polarization is said to be right-handed if the rotation resembles that of a right-handed screw advancing *against* the direction of propagation of the light, left-handed if the displacement is like that of a similarly advancing left-handed screw.

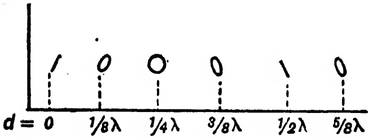


Fig. 624.

When light is totally reflected there is a phase difference between the vibrations respectively in and at right angles to the plane of incidence, so that this light is elliptically or circularly polarized. In ordinary reflection there is a slight elliptical polarization, which becomes very marked in the case of metallic reflection.

751. Production and Detection of Elliptically Polarized Light.—Circularly or elliptically polarized light cannot be detected by the unaided eye. If viewed through a Nicol prism no change in the intensity of circularly polarized light accompanies rotation of the prism, since a component of unchanging magnitude is transmitted. If the light is elliptically polarized there will be variations of intensity as the prism is rotated, the intensity being greatest when the principal section of the prism is parallel to the major axis of the ellipse (component amplitude of greatest magnitude) and a minimum when it is parallel to the minor axis. If circularly polarized light passes through a crystal producing a relative retardation of an odd number of quarter wave-lengths of a particular color the additional retardation between the components will cause the emergent light to be plane polarized in an aximuth which may be found by the analyzing Nicol prism. Such a crystal is called a *quarter-wave plate.* These plates can readily be prepared from thin sheets of mica.

It has been shown by R. A. Beth that elliptically polarized light carries with it angular momentum proportional to the area of the ellipse described by the light vibration (Fig. 624). When the light passes through a wave plate the ellipse is altered, and the change per second in angular momentum of the light can be measured as a moment of force that tends to turn the wave plate about an axis coinciding with the direction of propagation of the light.

752. Rotation of the Plane of Polarization.—If two Nicol prisms are set for extinction and a crystal of quartz cut with the face on which the light falls at right angles to the axis, or a solution of sugar or tartaric acid, is placed between them, the light will be restored. On turning the analyzer through an angle depending on the thickness of the substance and the color of the light, the light will again be extinguished. This shows that the plane of polarization has been

rotated through this angle. Substances producing this effect are said to be naturally optically active.

Some quartz crystals rotate the plane of polarization clockwise looking against the direction of propagation, and are called right-handed; others produce rotation in the opposite direction, and are called left-handed. These two classes of crystals can be distinguished by inspection of natural crystals on account of certain unsymmetrical facets which are differently placed in the two cases.

The rotation of the plane of polarization of light of the wave lengths corresponding to some Fraunhofer lines caused by a quartz plate of one mm. thickness is given below:

A	B	C	D	F	G	K
12.67°	15.75°	17.32°	21.70°	32.97°	42 60°	52.15°

As shown by these figures, the rotation varies very nearly inversely as the square of the wave-length.

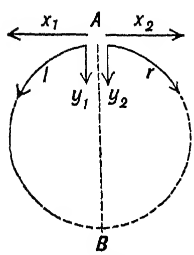


FIG 625

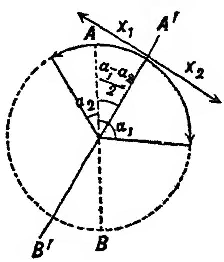


FIG. 626.

Fused quartz shows no double refraction or rotation. These effects are evidently due rather to the crystalline arrangement of the molecules than to their individual structure.

If light passes through a quartz prism so cut that the light is transmitted in the direction of the optic axis it is found that there is a very slight double refraction, so that spectral lines appear double. This shows that there are two waves traveling with slightly different velocities even along the optic axis. This is not generally true of uniaxial crystals, but only of those which rotate the plane of polarization. It is found that the two waves are circularly polarized in opposite directions, so that this is a case of circular double refraction. As first suggested by Fresnel, it appears that when light travels along the optic axis of quartz it is divided into two circularly polarized components, which travel with different velocities. These

on emergence recombine to form plane polarized light but in a different plane. This offers a simple explanation of the rotation.

If each circular displacement r and l is resolved into two linear displacements x and y, it is seen that when the two velocities of propagation of the two circular components are equal (Fig. 625) the two x components at any point in the medium are equal and opposite, leaving the two y components in the same direction to combine in a plane polarized beam, the vibrations of which are in the same direction as those of the original beam. But when the velocities of propagation are unequal (Fig. 626) the x components and the y components are respectively unequal. If, however, we refer displacements to an axis of reference shifted through an angle $(\alpha_1 - \alpha_2)/2$ with the original direction of vibration it will be seen that with reference to this axis the x displacements will cancel each other. This line $A'B'$ then represents the final direction of vibration and the rotation is $(\alpha_1 - \alpha_2)/2$.

In optically active crystals, like quartz, the two wave surfaces (Fig. 609) are not in contact even in the direction of the optic axis, and the complete theory shows that light traveling in any arbitrary direction is decomposed into two *elliptic* vibrations.

753. Rotation by Liquids. Saccharimetry.—A number of liquids, such as turpentine and the different sugars in solution, also cause rotation, that due to turpentine being left-handed, and that due to some sugars right-handed, to others left-handed. The vapors of such substances as turpentine also produce rotation. In such cases, as with quartz, there is circular double refraction, and the rotation is to be explained in the same way. In the case of liquids and vapors, however, the effect must be due to unsymmetrical structure of the molecule itself, as there is no crystalline structure, or if there is, the crystals are irregularly oriented. The amount of rotation varies approximately inversely as the square of the wave-length, and is proportional to the thickness of the medium, and also to the concentration in the case of solutions.

Hence if α is the rotation of light of a definite wave-length in passing a distance of l (decimeters) through a substance of density ρ gm/cm^3 or of percentage concentration p (grams per 100 cm^3 of solution), $\alpha = [\alpha]l\rho = [\alpha]lp/100$, where $[\alpha]$ is a constant for the substance called its *specific rotatory power*.

The rotatory power of the sugars is not affected by the presence of impurities. The percentage of sugar in a solution may therefore be determined by measuring the rotation with a polariscope. This is called *saccharimetry*. Most sugars rotate to the right, but levulose rotates to the left. In some cases the specific rotatory power varies

slightly with the concentration, and that of levulose is influenced by the temperature. The specific rotatory power for sodium light for some sugars at 20°C is given below (from Landolt, Optical Rotation). The positive sign indicates right-handed, the negative left-handed rotation, while p is the concentration.

Sucrose (cane sugar)	$+ 66.44° + 0.0087p$
Dextrose	$+ 52.50° + 0.0188p$
Levulose	$- 88.13° - 0.2583p$
Lactose (milk sugar)	$+ 52.53°$
Maltose (malt sugar)	$+ 140.4° - 0.0184p$

754. Rotation by Magnetic Field.—Faraday discovered that the plane of polarization of light passing through a refractive substance in a magnetic field is rotated if the light travels parallel to the force lines No effect is produced by a magnetic field on light waves in free space, and in general the effect increases with the refractive power of the substance, being especially marked in dense flint glass and carbon bisulphide and very feeble in the case of gases The rotation is usually proportional to the field intensity and to the thickness of the medium, and is in the direction of the positive electric current which might be used to produce the magnetic field. This is denoted right-handed rotation A very few substances show left-handed rotation. The effect varies with the wave-length, and the rotation produced by 1 cm. thickness in a field of one gauss (Verdet's constant) is: For water, 0.0131°; carbon bisulphide, 0 0435°; dense flint glass, 0.06° Enormous rotations are produced by thin films of iron or other magnetic material in a strong magnetic field.

In naturally active substances the direction of rotation is independent of the direction of propagation of the light, so that if a rotated beam is reflected its plane is turned back to the original position. In magnetically active substances the direction of rotation is reversed with reversal of the field, so that if the beam is reflected through the medium the rotation is doubled.

755. Magnetic Kerr Effect.—When a beam of plane-polarized light is reflected from a metallic surface a relative phase difference is introduced between components respectively in and at right angles to the plane of incidence, so that the reflected light is elliptically polarized, unless the incident light vibrates parallel or at right angles to the plane of incidence. Kerr found that if the light is reflected from the polished pole of an electromagnet it becomes slightly elliptically polarized, even under the conditions just mentioned.

756. Zeeman Effect.—Zeeman placed a bunsen flame colored with sodium between the poles of a powerful electromagnet. When the light from the source traveled either parallel or at right angles to the direction of the field, he observed a broadening of the spectral lines when the field was established. H. A Lorentz pointed out that such effects were in harmony with the *electron theory of radiation* proposed by him, and predicted that further investigation would show the radiation to be polarized by the field, either circularly or

plane, according to the direction in which it was viewed. Zeeman found this to be the case. In the simplest cases, when the light is viewed normally to the field, each spectral line is split into triplets, the vibrations in the central and undisplaced component being parallel to the force lines, those of the lateral and displaced components at right angles to the force lines. When the source is viewed parallel to the force lines single lines become doublets, the components being circularly polarized in opposite directions, and displaced on each side of the mean position of the line. In most cases the effects are much more complex, a large number of components being produced from single lines, but the simple case described above is explained by Lorentz's theory which assumes that the light waves are disturbances caused by rotations of their electrons about the atoms of the source, and that the motion of the electrons is modified by the magnetic field. The quantum theory of atomic structure (§758) leads to the *same* formula as does Lorentz's theory, in certain simple cases, and also explains completely the much more complicated Zeeman patterns shown by most spectral lines.

THE QUANTUM THEORY OF LIGHT

757. Corpuscular Properties of Light.—In our discussion of the behavior of light, we have assumed it to be a wave motion taking place in a hypothetical ether. We have made no further reference to the original idea of Newton (§624) that light is corpuscular in character. In recent years a body of evidence has accumulated which indicates that the energy of a beam of light is not spread out continuously over extended wave fronts, but is localized in individual elements of very small size. This has required a return to the corpuscular theory in a modified form.

The emission and absorption of light by atoms in descrete amounts called quanta (§729) gave the first indication that the light itself is discontinuous. More definite proof came from the study of the photoelectric effect (§544). The fact that the velocities of the ejected photoelectrons do not depend at all upon the intensity of the light, but only upon its frequency, could not be explained by the wave theory. Since the intensity determines the amplitude of the waves, greater intensity should mean greater amplitude and greater velocity of the photoelectrons. Instead, the photoelectrons have an energy determined by the frequency of the incident light. To explain this it is necessary to assume that light consists of separate quanta, each having an energy

$$E = h\nu.$$

The whole energy of one such quantum is communicated to each photoelectron. This shows that the quantum must be a concen-

trated bundle of energy of small size The special name *photon* has been given to this particle of light.

The use of rays in treating the problems of geometrical optics corresponds to the quantum conception of light, since the ray indicates the path of the photon. Now the shorter the wave-length of light, the less pronounced are the effects of diffraction, and the more nearly light "moves in straight lines" obeying the laws of geometrical optics. For this reason X-rays act much more like corpuscles than does ordinary light and it has been very difficult, experimentally, to prove that X-rays have the wave properties of ordinary light, although this has now been accomplished (§542). The existence of photons is strikingly shown for X-rays by a property known as the Compton effect. When X-rays are scattered by substances of low atomic weight, part of the scattered rays have a wave-length slightly longer than the incident rays. It is possible to give a quantitative explanation of this effect by assuming that photons are scattered by collision with electrons. Energy and momentum are conserved in the process, just as with colliding billiard balls, and the loss of energy of the scattered photon gives a lowering of frequency in exact agreement with experiment.

758. Wave Mechanics.—If the facts stated above indicate a radical departure from the classical theory of light, an even more remarkable departure from the classical mechanics of material particles has now been found necessary. Reasoning from certain analogies, it was first shown by de Broglie that a moving particle might have associated with it a system of waves, of wave-length given by

$$\lambda = \frac{h}{p}$$

Here h is Planck's constant, and p the momentum of the particle. For particles of ordinary mass the calculated wave-lengths are so small that the waves could never be detected. But in the case of electrons, the lightest known particles, we can obtain small values of p, and thus values of λ sufficiently large for these waves to be measurable. The wave-length of an electron having the velocity given by falling through a potential of 600 volts is about 0.5 Ångström unit, and thus comparable to that of X-rays. This led Davisson and Germer in 1937 to attempt to measure electron wave-lengths by diffraction from the surface of a crystal. A narrow beam of

electrons was directed against a nickel crystal, and the electrons reflected at various angles were measured. Maxima were found in certain directions corresponding to the various orders of a wavelength which agreed closely with that calculated by the above equation. This *diffraction of electrons* was later demonstrated in other ways, and it has even been possible to show the diffraction of a stream of atoms. Fig. 627 is a photograph of the diffraction pattern obtained when a narrow beam of electrons is passed through a film of molybdenum sulfide only 4×10^{-7} cm, or about seven atoms,

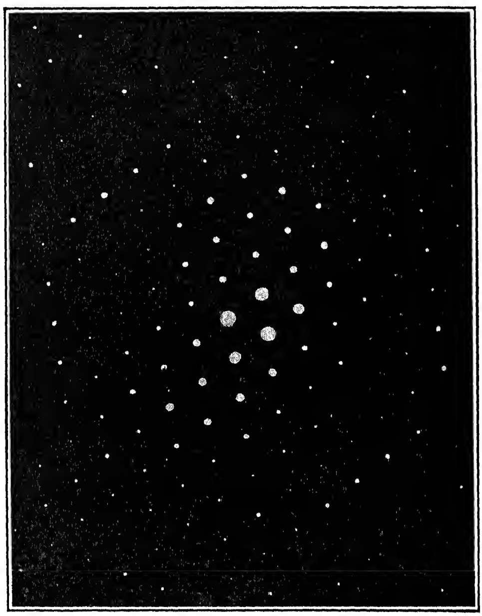


Fig. 627.

thick. The bright spots represent the different orders of the wavelength 0.07 Ångström unit (30,000 volts) produced by diffraction from the rows of atoms in the film. These rows constitute a "crossed grating," such as is obtained when two transmission gratings are superimposed with their rulings crossed. This gives a two-dimensional array of maxima instead of the one-dimensional array formed by an ordinary grating.

In dealing with very small particles such as electrons and atoms, it thus becomes necessary to use a mechanics based upon the

de Broglie waves. A theory of atomic structure and atomic processes known as the quantum mechanics or wave mechanics has now been developed from this standpoint. The formulation of this theory in 1926 was due to Schrödinger and Heisenberg The waves associated with the various particles are represented by mathematical equations, the solutions of which give the possible energy states of the system. Not only have many discrepancies that had arisen in the Bohr theory (§729) been successfully accounted for by the wave mechanics, but this theory is capable of dealing with a much wider variety of physical and chemical processes

As to the nature of the waves associated with a particle nothing is specified in the theory except that the square of their amplitude at any point in space gives the probability of finding the particle at that point. Hence they are frequently called *probability waves.* In the experiment of Davisson and Germer, the abundance of the electrons scattered at any angle is determined by the intensity of the probability waves diffracted at that angle. The distribution of electrons about the nucleus of an atom is represented as a system of stationary waves like those on a cord (§607) but more complex Just as with a rope, only certain modes of vibration are possible, and these represent the allowed states of the atom.

759. Heisenberg's Uncertainty Principle.—By the methods of wave mechanics we can calculate only the probability that a particle will be found in a given region of space. It is impossible to determine both the exact position and the exact momentum of the particle at any instant. This is the essence of a principle due to Heisenberg and Bohr, which fixes a theoretical limit to the accuracy with which these two quantities may be measured. Its basis lies in the fact that when we measure one of the two quantities, the other is altered by the process of measurement so that its value becomes uncertain. For example, the measurement of the position of an electron by any conceivable apparatus will introduce an uncertainty in its momentum. The uncertainty principle states that the product of the uncertainty in the position and the uncertainty of momentum can never be less than Planck's constant h. For heavy particles, this indeterminateness is entirely negligible because of the smallness of h, and Newtonian mechanics is adequate. For electrons and photons, however, it is of prime importance. Among other things it shows the impossibility of specifying exact orbits for the electron, as was done by Bohr.

760. Present Conception of the Nature of Light.—We have seen in §757 that light possesses the properties of particles, as well as those of waves. Similarly, in §758 it appeared that particles of ordinary matter also have wave properties. This dual character of both radiation and matter is an essential part of the wave mechanics. The light waves which produce interference and diffraction are merely the probability waves appropriate to photons, the particles of radiation. Thus light actually consists of discrete *particles.* At the same time, the paths followed by these particles are given by a *wave* theory,—that is they are given by the mathematical equations of ordinary wave motion. This accounts for the ability of light to produce interference patterns, etc. That the exact path followed by any individual photon can never be specified is shown by the uncertainty principle, but the wave theory can give us exact predictions of the paths followed on the average by a large number of photons, or the probability of a given path. As to why the particles should be governed by these waves of probability, the theory tells us nothing. Nevertheless the realization that both radiation and matter are probably corpuscular in character, but that both have wave-like properties, is one of the most important discoveries in the entire history of science.

PROBLEMS

1. A man is 5 feet 10 inches high. What is the shortest vertical plane mirror in which he can see his full-length image? *Ans.* 35 in.

Reflection.

2. Two plane mirrors are parallel to each other at a distance of 30 cm. Find the distance from each mirror of the three nearest images in each of an object between them and 10 cm. from one. *Ans.* 10, 50, 70; 20, 40, 80 cm.

3. A beam of light is reflected from a plane mirror revolving about a vertical axis ten times a second, and then falls on a plane vertical wall 20 meters from the mirror at the nearest point. What is the speed of the spot of light crossing the wall at this nearest point? How does the speed vary with the angle made by the reflected beam and the normal to the wall? *Ans.* 2513.6 m/sec; as $\sec^2\theta$.

4. With the sun at the zenith, a narrow beam of light entering through a small hole in a horizontal roof falls on the floor 20 ft. below. (*a*) At what speed, in inches per minute, is the spot of light moving across the floor? (*b*) If a mirror is placed flat on the floor, at what speed is the reflected spot moving across the ceiling? *Ans.* (*a*) 1.05 in./min. (*b*) 2.10 in./min.

5. Two plane mirrors make an angle of 60° with each other. Choose any arbitrary non-symmetrical position of an object between them and show by graphical construction the position of all possible images.

6. An object is 5 cm. high. It is desired to form a real image 2 cm. high, 100 cm. from the object. What is the character and focal length of the mirror which will do this? *Ans.* +47.7 cm.

7. A meter rod lies along the axis of a concave mirror of 20 cm. focal length, the zero end in contact with the mirror. Describe the image formed, and calculate the position of the first, fifth, tenth, twentieth, fortieth, and one-hundredth cm. marks, and the length of each division at these points (assuming the rod to be 2 cm. wide).

Ans. Virtual distances, (in cms.). 1.05, 6.67, 20, ∞; real, 40, 25. Lengths, 2.1, 2.67, 4, ∞, 2, 0.5.

8. A concave mirror has a radius of curvature of 50 cm. Find two positions in which an object may be placed in order to give an image four times as large. What is the position and character of the image in each case?

Ans. u = 31.25 cm., 18.75 cm.
v = +125 cm., −75 cm.
Image is real, virtual.

9. A convex mirror has a focal length of 25 cm. Calculate the position and the height of the image of an object 10 cm. high and 15 cm. in front of the mirror. *Ans.* −9.4 cm., 6.3 cm.

10. A small light is situated on the axis of a concave mirror of 20 cm. focal length at a distance of one meter from the mirror. A plane mirror is placed 110 cm. from the concave mirror so that the source of light lies between the two mirrors and the plane of the plane mirror is perpendicular to the axis of the concave mirror. Find (*a*) the distances behind the plane mirror of the first three images, (*b*) the relative sizes of the image.

Ans. (*a*) 10, 85, and 86 cm.
(*b*) $I = O$, $O/4$, and $O/5$.

11. The sun has an angular magnitude of 32′. What is the size of the solar image formed by a concave mirror of 50 ft. focal length? *Ans.* 5.58 in.

Refraction. **12.** In order to measure the index of refraction of a certain crystal, 12.62 mm. thick, a microscope is first focused on a small scratch on the upper surface and then on one on the lower surface as viewed through the crystal. The microscope is lowered 5.43 mm. between these two settings. What is the index of refraction? *Ans.* 2.33.

13. A layer of ether (n = 1.36) 2 cm. deep floats on water (n = 1.33) 3 cm. deep. What is the apparent distance of the bottom of the vessel below the surface? *Ans.* 3.72 cm.

14. An object is viewed through a cube of glass (n = 1.55) 10 cm. thick, in a direction at an angle of 60° with the normal to the glass surface. What is the lateral displacement of the image? *Ans.* 5.29 cm.

15. Derive a general expression for the lateral displacement of a light ray passing through a plate with parallel sides, in terms of the thickness of the plate, the angle of incidence, and the angle of refraction. *Ans.* $\dfrac{d \sin (i - r)}{\cos r}$,

Lenses. **16.** The object glass of the Yerkes telescope has a focal length of 62 feet. Assuming that the angular diameter of the sun is 32 minutes, find the diameter in inches of the image of the sun produced by the objective. *Ans.* 6.9 inches.

17. The focal length of a double convex lens is one foot. (*a*) Where must an object be placed so that the image will be 18 inches from the lens and real? (*b*) 18 inches from the lens and virtual? (*c*) If the object is 2 inches high, what will be the size of the image in each case? Draw a diagram for each of the two cases. *Ans.* (*a*) 36 inches from lens.
(*b*) 7.2 inches from lens.
(*c*) 1 inch, 5 inches.

18. A convex lens 25 cm. from a candle-flame 5 cm. high forms an image of the latter on a screen. When the lens is moved 25 cm. further from the candle an image is again formed on the screen. Calculate the focal length of the lens, the distance of the screen from the candle, and the size of the two images. *Ans.* 16.67; 75; 10, 2.5 cm.

19. A projecting lantern is to produce a magnification of 50 diameters at a distance of 40 feet from the objective. Find the distance of the lens from the lantern slide, and the equivalent focal length of the objective lens.
Ans. 9.6 inches.
9.4 inches.

20. A candle flame 100 cm. from a convex lens of focal length 90 cm. is displaced 2 cm. away from the lens at the rate of 1 cm. per second. What is the displacement and the average velocity of its image?

Ans. 135 cm. toward lens; 67.5 $\frac{\text{cm.}}{\text{sec.}}$

21. A convex lens (n = 1.54) has a focal length of 40 cm. in air. What is the focal length in water (n = 1.33)? *Ans.* 136.8 cm.

22. A plano-convex lens, made of glass of index of refraction 1.55, is to be immersed in water of index 1.33. What must be the radius of the curved surface of the lens in order that the resulting focal length of the lens in water may be 50 inches? *Ans.* 8.25 inches.

23. Two convex lenses, each having a focal length of 12 cm., are placed 8 cm. apart. An object is placed 3 cm. in front of the first lens. Where is the image? *Ans.* At infinity.

24. Two convex lenses of focal lengths 20 and 30 cm. are 10 cm. apart. Calculate the position and length of the image of an object 2 cm. long 100 cm. in front of the first lens. *Ans.* 10 cm. beyond second lens; length 0.33 cm.

25. Replace the first lens in the above problem by a concave lens of the same focal length and determine the position and magnitude of the image.
Ans. 240 cm. to left of second lens; 3 cm.

26. The eye-piece of a telescope has a focal length of one cm. When focused on a star, the final image being at infinity, the distance between the eye-piece and objective is 51 cm. To see a certain terrestrial object clearly, with the final image again at infinity, the eye-piece must be drawn out 0.2 cm. What is the distance of the object from the telescope? *Ans.* 125.5 m.

27. In the above example, if the object referred to is a tree 10 feet high, what is the size of the image formed by the object lens? What is the angular magnitude of the image formed by the eye-piece? *Ans.* 0.04 feet.
62° 28′.

28. A double convex lens with faces having a radius of curvature of 30 cm. gives a real image at a distance of 60 cm. of an object 40 cm. away. What is its focal length? Its index of refraction? *Ans.* 24 cm.; 1.625.

29. The index of refraction of a certain kind of glass is 1.5 for red light and 1.6 for violet light. A plano-convex lens is constructed of this glass, the radius of the curved face being 25 cm. Find (*a*) the focal length for red and for violet light, (*b*) the linear separation of the red and violet portions of the image of an object 30 cm. from the lens.

Ans. (*a*) +41.66 cm., +50.0 cm.
(*b*) 32.14 cm.

30. An achromatic lens is to be made of a combination of a crown glass double convex lens ($n_D = 1.51$, $n_F = 1.52$) and a plano-concave flint glass lens ($n_D = 1.64$, $n_F = 1.66$), the adjacent faces to fit together and the focal length to be 50 cm. Calculate the radii of curvature of the faces.

Ans. $r_1 = r_2 = -r_3 = 19$ cm.; $r_4 = \infty$.

31. Light from an object 2 inches high passes through a convex lens of 15 inches focal length, situated 30 inches from the object, and then falls on a concave mirror of 30 inches focal length placed 20 inches behind the lens. What is the position, size, and character of the final image?

Ans. $7\frac{1}{2}$ inches in front of mirror,
$1\frac{1}{2}$ inches high, real.

32. In a given microscope the equivalent focal length of the objective is 8 mm. and of the eye-piece is 20 mm. The distance between the objective and eye-piece is 150 mm. (*a*) If the final image is formed at infinity, what is the magnifying power of the microscope? (*b*) How much will this be changed by placing the final image at the distance of most distinct vision (10 inches)?

Ans. (*a*) 193.67.
(*b*) Changed to 211.64.

33. A concave lens of 5 cm. focal length is placed at such a distance behind a convex lens of 25 cm. focal length that the final image of an object 4 meters in front of the convex lens is formed 60 cm. behind the concave lens. What is the distance between the two lenses? What is such a combination of lenses called?

Ans. 22.05 cm. Telephoto lens.

34. A telephoto camera has a convex lens of 20 cm. focal length, and, 15.5 cm. behind it, a concave lens of 5 cm. focal length. What is (*a*) the back focal length of the lens (i.e. the actual distance behind the convex lens of the image when an object is at infinity)? (*b*) the focal length of a convex lens which alone would give an image of the same size as the above telephoto combination?

Ans. (*a*) 60.5 cm.
(*b*) 200 cm.

35. The focal length of a convex lens is 6 inches. Behind it is placed a concave lens of 2 inches focal length, at such a distance that this concave lens forms a virtual image 10 inches from itself, of an object 5 ft. high and 30 ft. in front of the convex lens. Find (*a*) the distance between the two lenses, (*b*) the size of the virtual image, (*c*) the magnifying power of the combination. (*d*) What is such a combination of lenses called?

Ans. (*a*) 3.60 inches.
(*b*) 1.01 inches.
(*c*) 2.44.
(*d*) Opera glass.

36. A double convex lens is constructed of glass of index of refraction 1.516 and radii of curvature 12 inches. One of the surfaces is silvered so that it acts

as a mirror and reflects back into the glass light which enters the lens from the other surface. Where will such a combination bring to a focus a source of light placed 50 inches from it in such a manner that the light first strikes the non-silvered surface? *Ans.* 3.14 inches from lens.

Photometry. **37.** Two sources of light have candle powers of 20 and 80 respectively. (*a*) If they are 100 cm. apart, at what two points on a straight line passing through the two sources can a screen be placed in order to be equally illuminated by the two? (*b*) Are there other points, not on this line, where the illumination is equal?

Ans. (*a*) $33\frac{1}{3}$ cm. inside, or 100 cm. outside the fainter source.
(*b*) Yes.

38. A candle is placed 10 cm. in front of a concave mirror of 20 cm. focal length (assumed to be a perfect reflector). What is the illumination on a screen 100 cm. from the candle along the mirror axis, as compared with I, that due to the candle alone? *Ans.* 3.37 I.

39. Solve the above problem after substituting a convex mirror of the same focal length for the concave mirror. *Ans.* 1.35 I.

40. A 60 watt tungsten light with a rated efficiency of 1.2 watts per candle illuminates a surface 8 feet distant. What is (*a*) the mean spherical candle-power of the light? (*b*) the total luminous flux (in lumens)? (*c*) the lumens per sq. meter (lux) on the above surface? (*d*) the lumens per sq. ft. (ft. candles)?

Ans. (*a*) 50 c.p.
(*b*) 628.3.
(*c*) 8.42.
(*d*) 0.782.

41. The full moon at the zenith gives an illumination of 0.014 lumens per sq. ft. (*a*) At what distance will a 500 candle power lamp give the same illumination? (*b*) If the illumination due to a sixth magnitude star is equal to that of 1 candle at 7 miles, how many such stars would give an illumination equal to that of the full moon at the zenith? *Ans.* (*a*) 189 ft.
(*b*) 1.914×10^7.

42. (*a*) How far above a table should a 32 candle power lamp be placed in order to give a direct illumination of 4 foot-candles? (*b*) How many lumens fall on a square meter of this table top? (*c*) The sun at the zenith gives an illumination of 9000 lumens per sq. ft. At what angle to the vertical would the table top have to be inclined in order to receive directly from the sun the same intensity as from the electric light? *Ans.* (*a*) 2.83 ft.
(*b*) 43
(*c*) 1.525′ of arc.

Dispersion. **43.** A 60° prism is made of glass of index of refraction 1.55 for sodium light. What is the correct angle of incidence and the total angular deviation, when the prism is in the position of minimum deviation, in air? What are these two angles when the prism is in water of index of refraction 1.33? *Ans.* $i = 50°48'$
$D = 41°36'$.
$i' = 35°37'$.
$D' = 11°14'$.

44. A 60° prism has an index of refraction of 1.62 for the D lines and 1.63 for the F line. If white light is incident at an angle of 45 degrees, what are the respective angles of emergence for these two colors?

Ans. 65° 19′: 66° 40′.

45. In the above case, what is the angle of minimum deviation for each color? If the spectrometer telescope has a focal length of 30 cm., what is the length of the spectrum between D and F when the prism is set for minimum deviation for the D lines? *Ans.* D, 48° 12′; F, 49° 10′; .51 cm.

46. Light incident internally on the surface of a glass prism at an angle of 56° is totally reflected from a drop of liquid in contact with the glass. If the index of refraction of the latter is 1.62 for sodium light, what is the index of refraction of the liquid? *Ans.* 1.343.

Interference. **47.** In a system of Newton's rings due to a convex lens resting on a plane surface the 25th ring is 1 cm. from the center, when sodium light is used. What is the thickness of the air film at that point, and what is the radius of curvature of the lens?

Ans. 0.00721 mm.; 6.93 meters.

48. If the air film is replaced by water in the above example, what will be the distance of the 25th ring from the center? *Ans.* 0.87 cm.

49. Light from a narrow slit passes through two parallel slits 0.2 mm. apart. The interference bands on a screen 100 cm. away are 2.95 mm. apart. What is the wave-length of the light? *Ans.* 0.00059 mm.

50. The angles of a Fresnel biprism are 10′ and the index of refraction 1.62. What is the distance between the two images of a slit 20 cm. from the prism? What is the width of the interference bands of sodium light formed on a screen 50 cm. beyond the prism? What is their width if light of the wave-length of the F line is used? *Ans.* 0.724 mm.; 0.57 mm.; 0.47 mm.

51. A film of glass of index of refraction 1.54 is introduced in one of the interfering beams of Michelson interferometer, and causes a displacement of 20 fringes of sodium light across the field. What is the thickness of the film?

Ans. 0.0109 mm.

52. The D lines in the spectrum of the second order formed by a Rowland concave grating of 15 feet radius of curvature are 315 cm. from the slit. What is the distance between rulings? *Ans.* 0.00171 mm.

Diffraction. **53.** The central maximum of the diffraction bands of sodium light produced by a narrow slit on a screen at a distance of 100 cm. is 2 mm. wide. How wide are the other maxima and the slit?

Ans. 1 mm.; 0.589 mm.

54. Two narrow slits 0.1 mm. apart are illuminated by sodium light. What must be the diameter of a lens 5 meters away to clearly resolve the images of the two slits? *Ans.* 3.54 cm.

55. In the above case, at what distance will the same lens clearly resolve the images of the slits if they are illuminated by light of wave-length corresponding to that of the F line? *Ans.* 6.05 m.

Polarization. **56.** Determine the angle of incidence, on a smooth water surface, of sodium light such that the reflected beam shall contain a maximum amount of plane polarized light. *Ans.* 53°7′.

57. A 30° calcite prism is cut with its refracting edge parallel to the optic axis. From the indices of refraction as given in the text find the angle of minimum deviation for the ordinary and for the extraordinary ray for sodium light.

Ans. $D_o = 20°50'$.
$D_e = 15°15'$.

58. It is desired to obtain plane polarized light of $\frac{1}{8}$ the intensity of the unpolarized incident light, by using two Nicol prisms. (*a*) What must be the angle between the short diagonals of the two Nicols in order to accomplish this losses due to ordinary reflection and absorption being neglected? (*b*) Considering the variation of intensity of the transmitted light, as the angle between two Nicols is varied, at what angle is the rate of variation of intensity with angle an algebraic minimum? *Ans.* (*a*) 60°.
(*b*) 45°.

59. Plane polarized light falls normally on a plate of quartz with faces parallel to the axis. If the vibrations of the incident light are at an angle of 30° with the principal section, calculate the relative intensities of the transmitted ordinary and extraordinary rays. *Ans.* 0.25, 0.75.

60. In the above case, if the crystal is 1 mm. thick, what is the difference of phase upon emergence of the ordinary and extraordinary rays of sodium light (§743)? *Ans.* 30.90π radians.

61. A crystal of Iceland spar cut with faces parallel to the axis is 2 cm. thick. How far below the upper surface are the ordinary and extraordinary images of a pencil mark on the lower face? *Ans.* 1.206, 1.346 cms.

62. (*a*) If a plate of quartz is cut with the faces perpendicular to the optic axis, how thick must it be to rotate through 90°, the plane of polarization of sodium light falling normally on it? (*b*) What is the necessary thickness of a 10% malt sugar solution (i.e. 10 grams per 100 cm.^{3}) in order to cause the same rotation? *Ans.* (*a*) 4.15 mm.
(*b*) 64.2 cm.

63. Through how many degrees will a column 20 cm. long of a 10 per cent. solution of cane sugar rotate the plane of polarization of sodium light?
Ans. 13.30°.

Spectrum.

64. On mapping the spectral intensity curve of an incandescent source it is found that the maximum intensity is at a wave-length 12,000 Ångström units. What is the temperature of the source? *Ans.* 2403° abs.

65. The displacement of the F line of hydrogen (wave-length 4861 Ångström units) in the spectrum of a star is 0.1 of an Ångström unit toward the violet. What are the direction of motion and the velocity of the star in the line of sight? *Ans.* $6.2\,\dfrac{\text{km.}}{\text{sec.}}$ towards earth.

Logarithms of Numbers from 1 to 1000.

	1	2	3	4	5	6	7	8	9
00	0043	0086	0128	0170	0212	0253	0294	0334	0374
14	0453	0492	0531	0569	0607	0645	0682	0719	0755
92	0828	0864	0899	0934	0969	1004	1038	1072	1106
39	1173	1206	1239	1271	1303	1335	1367	1399	1430
61	1492	1523	1553	1584	1614	1644	1673	1703	1732
61	1790	1818	1847	1875	1903	1931	1959	1987	2014
41	2068	2095	2122	2148	2175	2201	2227	2253	2279
04	2330	2355	2380	2405	2430	2455	2480	2504	2529
53	2577	2601	2625	2648	2672	2695	2718	2742	2765
88	2810	2833	2856	2878	2900	2923	2945	2967	2989
10	3032	3054	3075	3096	3118	3139	3160	3181	3201
22	3243	3263	3284	3304	3324	3345	3365	3385	3404
24	3444	3464	3483	3502	3522	3541	3560	3579	3598
17	3636	3655	3674	3692	3711	3729	3747	3766	3784
02	3820	3838	3856	3874	3892	3909	3927	3945	3962
979	3997	4014	4031	4048	4065	4082	4099	4116	4133
150	4166	4183	4200	4216	4232	4249	4265	4281	4298
314	4330	4346	4362	4378	4393	4409	4425	4440	4456
472	4487	4502	4518	4533	4548	4564	4579	4594	4609
624	4639	4654	4669	4683	4698	4713	4728	4742	4757
771	4786	4800	4814	4829	4843	4857	4871	4886	4900
914	4928	4942	4955	4969	4983	4997	5011	5024	5038
051	5065	5079	5092	5105	5119	5132	5145	5159	5172
185	5198	5211	5224	5237	5250	5263	5276	5289	5302
315	5328	5340	5353	5366	5378	5391	5403	5416	5428
5441	5453	5465	5478	5490	5502	5514	5527	5539	5551
5563	5575	5587	5599	5611	5623	5635	5647	5658	5670
5682	5694	5705	5717	5729	5740	5752	5763	5775	5786
5798	5809	5821	5832	5843	5855	5866	5877	5888	5899
5911	5922	5933	5944	5955	5966	5977	5988	5999	6010
6021	6031	6042	6053	6064	6075	6085	6096	6107	6117
6128	6138	6149	6160	6170	6180	6191	6201	6212	6222
6232	6243	6253	6263	6274	6284	6294	6304	6314	6325
6335	6345	6355	6365	6375	6385	6395	6405	6415	6425
6435	6444	6454	6464	6474	6484	6493	6503	6513	6522
6532	6542	6551	6561	6571	6580	6590	6599	6609	6618
6628	6637	6646	6656	6665	6675	6684	6693	6702	6712
6721	6730	6739	6749	6758	6767	6776	6785	6794	6803
6812	6821	6830	6839	6848	6857	6866	6875	6884	6893
6902	6911	6920	6928	6937	6946	6955	6964	6972	6981
6990	6998	7007	7016	7024	7033	7042	7050	7059	7067
7076	7084	7093	7101	7110	7118	7126	7135	7143	7152
7160	7168	7177	7185	7193	7202	7210	7218	7226	7235
7243	7251	7259	7267	7275	7284	7292	7300	7308	7316
7324	7332	7340	7348	7356	7364	7372	7380	7388	7396
0	1	2	3	4	5	6	7	8	9

Logarithms of Numbers from 1 to 1000.

No.	0	1	2	3	4	5	6	7	8	9
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996
No.	0	1	2	3	4	5	6	7	8	9

°	Sin	Tan	Cot	Cos	°
0	0.0000	0.0000	∞	1.0000	90
1	0.0175	0.0175	57.2900	0.9998	89
2	0.0349	0.0349	28.6363	0.9994	88
3	0.0523	0.0524	19.0811	0.9986	87
4	0.0698	0.0699	14.3007	0.9976	86
5	0.0872	0.0875	11.4301	0.9962	85
6	0.1045	0.1051	9.5144	0.9945	84
7	0.1219	0.1228	8.1443	0.9925	83
8	0.1392	0.1405	7.1154	0.9903	82
9	0.1564	0.1584	6.3138	0.9877	81
10	0.1736	0.1763	5.6713	0.9848	80
11	0.1908	0.1944	5.1446	0.9816	79
12	0.2079	0.2126	4.7046	0.9781	78
13	0.2250	0.2309	4.3315	0.9744	77
14	0.2419	0.2493	4.0108	0.9703	76
15	0.2588	0.2679	3.7321	0.9659	75
16	0.2756	0.2867	3.4874	0.9613	74
17	0.2924	0.3057	3.2709	0.9563	73
18	0.3090	0.3249	3.0777	0.9511	72
19	0.3256	0.3443	2.9042	0.9455	71
20	0.3420	0.3640	2.7475	0.9397	70
21	0.3584	0.3839	2.6051	0.9336	69
22	0.3746	0.4040	2.4751	0.9272	68
23	0.3907	0.4245	2.3559	0.9205	67
24	0.4067	0.4452	2.2460	0.9135	66
25	0.4226	0.4663	2.1445	0.9063	65
26	0.4384	0.4877	2.0503	0.8988	64
27	0.4540	0.5095	1.9626	0.8910	63
28	0.4695	0.5317	1.8807	0.8829	62
29	0.4848	0.5543	1.8040	0.8746	61
30	0.5000	0.5774	1.7321	0.8660	60
31	0.5150	0.6009	1.6643	0.8572	59
32	0.5299	0.6249	1.6003	0.8480	58
33	0.5446	0.6494	1.5399	0.8387	57
34	0.5592	0.6745	1.4826	0.8290	56
35	0.5736	0.7002	1.4281	0.8192	55
36	0.5878	0.7265	1.3764	0.8090	54
37	0.6018	0.7536	1.3270	0.7986	53
38	0.6157	0.7813	1.2799	0.7880	52
39	0.6293	0.8098	1.2349	0.7771	51
40	0.6428	0.8391	1.1918	0.7660	50
41	0.6561	0.8693	1.1504	0.7547	49
42	0.6691	0.9004	1.1106	0.7431	48
43	0.6820	0.9325	1.0724	0.7314	47
44	0.6947	0.9657	1.0355	0.7193	46
45	0.7071	1.0000	1.0000	0.7071	45
°	Cos	Cot	Tan	Sin	°

PHYSICAL CONSTANTS

Å	Ångstrom unit	1 Å = 10^{-8} cm = 10^{-4} micron
c	Velocity of light	2 998 × 10^{10} cm /sec.
d_{air}	Density of dry air N T P	0 001293 gram/cm.^{3}
d_{H_2O}	Density of water	1 gram/cm ^{3} = 62.4 lb./ft.^{3}
d_{Hg}	Density of mercury	13.596 gram/cm.^{3}
e	Electronic charge	4 803 × 10^{-10} e s u = 1 603 × 10^{-20} e.m.u.
F	Faraday	9.65 × 10^4 coulomb
G	Gravitation constant	6 66 × 10^{-8} dyne cm.^{2}/gram^2
g	Acceleration of gravity at 45°	980 6 cm./sec ^{2} = 32.2 ft./sec.^{2}
H	Heat	1 B.T.U. = 252 cal.
H_f	Heat of fusion of ice	79.6 cal./gm.
H_v	Heat of vaporization of water	539 cal /gm.
h	Planck's constant	6.55 × 10^{-27} erg sec.
J	Mechanical equivalent of heat	4 185 × 10^7 ergs/cal. = 778 ft. lbs./B.T.U.
k	Boltzmann's constant	1.371 × 10^{-16} erg/°C.
l	Length	1 meter × 39.37 inches 1 inch = 2 54 cm.
m	Mass	1 kg = 2.205 lb. 1 lb. = 453.6 grams
m_0	Mass of O^{16} atom	2.64 × 10^{-23} gram
m_e	Mass of an electron	9.10 × 10^{-28} gm.
N	Avogadro's number	6.06 × 10^{23} per gr. mol.
P	Power	1 H.P. = 550 ft lb./sec. = 745 8 watt
P_0	Standard atmosphere	76 cm. Hg = 1.013 × 10^6 dynes/cm.^{2} = 14.7 lb./in.^{2}
R	Gas constant	8.31 × 10^7 ergs/°C. gr. mol.
s	Stefan's radiation constant	5.7 × 10^{-5} erg/cm.^{2} sec.
T_0	Ice point	0°C. = 32°F. = 273.18°K.
T_{100}	Steam point	100°C. = 212°F. = 373.18°K.
V_0	Volume of gr. mol of gas at N T.P.	22.41 × 10^3 cm.^{3}
v_s	Velocity of sound in dry air at 0°C.	332 m /sec. = 1070 ft./sec.

For electrical units see Table on page 440.

ADDITIONAL PROBLEMS

Mechanics and Properties of Matter

Velocity and Acceleration.

1. Two automobiles moving at 20 miles per hour and 30 miles per hour pass at a cross road. What is their speed of separation (the difference of their velocities)?
2. In what direction must a boat be rowed at 5 miles per hour to go directly across a river flowing at a rate of 3 miles per hour?
3. If the acceleration with which a car can be started is reduced by 50 per cent by a bad road, in what proportion is the time required to get up a certain speed increased?
4. How long does it take to bring an automobile travelling at 30 miles per hour to rest in a distance of 40 feet?
5. The speed of a train increases from 20 miles per hour to 30 miles per hour in a distance of 1000 ft. Calculate the acceleration, assumed constant.
6. With what velocity must a body be thrown vertically upward to return to the starting point in 8 seconds?
7. On June 30, 1937, Lieut. M. J. Adam of the Royal Air Force reached a height of 53,937 ft. If he had dropped a bit of lead from the airplane when it was travelling at 100 miles per hour horizontally, how long would it have taken to reach the ground, and how far would it have travelled horizontally, *if* there had been no air resistance?
8. In a laboratory experiment a ball projected horizontally fell a vertical distance $y = 12.5$ cm. in travelling a distance $x = 25$ cm. horizontally. At what point in its fall were x and y equal and what was then the direction of motion?
9. A baseball is thrown with a speed of 20 meters per sec. at an elevation of 60°. How high will it rise and how far will it travel horizontally?
10. What is the speed and what is the angle of projection of a baseball that rises to a height of 64 ft. and travels 128 ft. horizontally?
11. A body is projected at an elevation of 45° and remains 5 seconds in the air. What was the initial speed and what was the horizontal distance travelled?
12. In what direction should a jet of water leave the nozzle with a speed of 32 ft. per sec. to fall on a flower-bed 32 ft. away and 8 ft. below?
13. What is the central acceleration of an automobile that is going around a curve of 100 ft. radius at 30 miles per hour?
14. Calculate in cm. per sec.^{2} the central acceleration of a point on the earth's equator, the radius of which is about 6400 km.

Force and Mass.

15. Two forces of equal magnitude act at a point. Draw a vector diagram for finding the angle between their directions if their resultant is equal to each in magnitude.

16. Forces proportional in magnitude to 4, 5, 6, acting at a point are in equilibrium. Calculate the angles between their directions.
17. A body of 1 kg. mass, starting from rest, moves a distance of 200 cm. with a constant acceleration in 5 sec. What force acts on it?
18. A body that weighs 10 lbs. is pulled upward by a force equivalent to 12 lbs. weight. With what acceleration does it start?
19. A skater who weighs 160 lbs. comes to rest in gliding 40 yards in 10 seconds. What is the force of friction?
20. A shot that weighs 100 lbs. is brought to rest by impact in 0.01 sec. What average force does it exert?
21. What is the acceleration with which an elevator starts if a mass of 100 gms. on a spring balance on the elevator seems to weigh 105 gms.?
22. What is the tension in a rope carrying a bucket which weighs 100 lbs. and descends with an acceleration of .5 ft. per sec.^{2}?
23. A body that weighs 100 lbs. is on a smooth plane inclined at 30° to the horizontal and is pushed up by a horizontal force equal to the weight of 60 lbs. What is the acceleration?
24. A body slides down a smooth plane inclined at 30° to the horizontal. What is its speed at the end of the third second and how far does it travel in the fourth second?
25. A toboggan slides with constant acceleration down a toboggan slide 100 yards long and inclined at 10° in 13 seconds. What final velocity does it acquire and by how much is this velocity decreased by the friction?
26. If a boy who weighs 120 lbs. jumps horizontally from a small boat that weighs 500 lbs. what is the ratio of his velocity relative to the shore to his velocity relative to the boat?

Work and Energy.

27. Calculate the number of foot-pounds in a meter-kilogram, the engineering unit of work in countries using the metric system.
28. A body of 100 g. mass revolves in a circle of 25 cm. radius making 120 revolutions per minute. How large is the force toward the center, and how much work does it do in a minute?
29. A man climbs a mountain 2000 feet high in 2 hours. At what additional rate would he have to work if he carried a knapsack weighing 40 lbs?
30. How much work is done in drawing a load of 500 lbs. 10 feet up an incline of 30°, if the resistance of friction is 100 lbs?
31. Compare the energy required to give an automobile a speed of 30 miles per hour with that required to raise it 30 feet.
32. A locomotive can just draw a train of 5000 tons up an incline of one degree. In what distance could the locomotive give the train a speed of 10 miles per hour on a level track?
33. A pendulum consists of a bob of 250 g. mass attached to a light wire. When it is drawn aside 30° and released, with what energy does it pass through the vertical?
34. A hammer of 4 kg. mass moving at 4 cm. per sec. drives a nail 1 cm. into soft wood. What is the average resistance? If the resistance is proportional to the depth, how much further will the nail be driven by a second similar blow?

35. Bodies of 50 gm. and 300 gm. weight hang by a cord that passes over a light frictionless pulley. What is the loss of potential energy when each has moved 10 cms. from rest and at what speed are they moving?

36. A bullet that weighs 10 gms. is fired with a speed of 800 meters per sec. into a block of 10 kg. hanging from cords 2 m. long. Through what angle does the pendulum swing out?

Rotation.

37. A fan is revolving at the rate of 20 revolutions per second and when the power is shut off it comes to rest in 5 seconds. How many revolutions does it make and what is the angular acceleration?

38. An automobile is travelling at 30 miles per hour. The outside diameter of a wheel is 28 inches. What is the velocity of a point at the top of a wheel relatively to the ground and relatively to the center?

39. How much does the speed of the top of the Empire State building (1248 ft.) exceed that at the bottom (radius of earth 4000 miles, latitude of New York 41° N.)?

Center of Mass.

40. Equal masses are placed at the corners of a triangle of sides 30, 40, 50 cms. Where is their center of mass?

41. Where is the center of mass of a uniform disk having the form of a square of 10 cm. side extended into an equilateral triangle on one side?

42. Two homogeneous cylinders of the same material each 20 cm. long and of radii 4 cm. and 3 cm. are joined end to end coaxially. Where is the center of mass of the whole?

43. A circular table 3 ft. in diameter and 2 inches thick has three legs each 2×2 in. and 2 ft. long. Where is the center of mass of the table?

44. From an equilateral triangle of 20 cm. side a smaller equilateral triangle of half the height is cut. Where is the center of gravity of the remaining trapezoid?

Moments.

45. What is the least force that, applied to a cube of 3 ft. edge and 800 lbs. weight, will tip it around an edge, if it is kept from slipping?

46. A circular disk of 20 cm. radius and 1 kg. mass is in rotation making 300 revolutions per minute. What is its kinetic energy?

47. A silver half dollar weighs 13.4 g. and its radius is 1.5 cm. Calculate its rotational inertia about an axis through its center parallel to each face.

48. Calculate the linear acceleration of a hoop rolling down a plane inclined at 20° to the horizontal.

49. To the lower end of a vertical rod that weighs 500 g. and is 100 cm. long a sphere of 5 cm. radius that weighs 1000 g. is attached. Calculate the radius of gyration about the upper end.

50. If the apparatus of the preceding problem is free to swing as a pendulum about its upper end and a horizontal force equal to the weight of 200 grams is applied in a line through the center of the sphere, with what angular acceleration will the pendulum start?

51. A cylinder is mounted with its axis horizontal so that it can rotate with negligible friction. A thin cord wrapped around the cylinder carries a

block of mass equal to that of the cylinder. With what acceleration will the block descend when released?

52. You have probably seen a photograph of a tall chimney falling and breaking into two parts as it fell. Why did it break?

Parallel Forces.

53. Forces, each of 3 pounds, act at points A, B, C in a line ABC in a direction perpendicular to ABC, the first two being in the same direction and the third in the opposite direction. If $AB = BC = 2$ ft. at what point in ABC does the resultant act?

54. A man and a boy carry a uniform beam 10 ft. long that weighs 200 pounds. If the boy holds one end, where must the man take hold to carry three-fourths of the weight?

55. A plank 20 ft. long weighs 100 pounds, is supported at the ends, and carries a weight of 80 lbs. 5 ft. from one end. Calculate the pressures on the supports.

56. Four parallel forces, of magnitudes 3, 6, 9, 12 kg., act at the corners of a square of 50 cm. side, the first three in the same direction and the fourth in the opposite direction. Find the line of action of their resultant.

Equilibrium.

57. A rod is suspended by two cords at its ends and makes an angle of 45° with the vertical. If the lower cord is horizontal, what is the direction of the upper one?

58. A uniform ladder leans against a wall and the friction is negligible. If the pressure against the wall is equal to one-third the weight of the ladder, at what angle is the ladder inclined to the vertical?

59. The hinges on a door that weighs 250 pounds are 6 ft. apart. If the weight is all on the lower hinge, what is the resultant force at this hinge?

60. A board that weighs 100 pounds is kept at 30° to the horizontal by a cord attached to the upper end at right angles to the board. What is the tension on the cord and what is the total force exerted on the board at the lower end?

Periodic Motions.

61. What is the maximum acceleration of an end of a tuning-fork that makes 435 complete vibrations per second if the amplitude is 2 mm.?

62. The period of a torsion pendulum is 5 sec. What is the angular acceleration at the moment when the angular displacement is 45°?

63. The value of g is 980.07 cm./sec.^{2} at Washington and 980.23 cm./sec.^{2} at New York. What is the ratio of the periods of the same pendulum at the two places?

64. If the equivalent length of a clock pendulum increases by 0.01 % because of thermal expansion, how much does the clock lose per day?

65. What is the period of vibration of a uniform rod 150 cm. long, suspended at one end? About what other points will the rod vibrate with the same period?

66. A uniform disk, 30 cm. in diameter hangs from a string of 30 cm. length attached to a point on the rim. What is its period of vibration in its own plane?

67. To find the shear modulus of the material of a wire from its constant of torsion, a circular disk of 5 cm. radius and 1000 gms. mass is suspended

from the wire with its plane horizontal. If it makes 12 torsional vibrations per minute, what is the constant of torsion of the wire?

68. To find the rotational inertia of an irregular disk it is used as the bob of a torsion pendulum with its plane horizontal, and the pendulum makes 10 vibrations per minute. A circular plate of 5 cm. radius and 250 gms. mass is then placed on the disk and the number of vibrations per minute is reduced to 5. Calculate the rotational inertia of the disk.

Friction.

69. What force is required to start the motion of a block that weighs 25 lbs. on a horizontal surface if the coefficient of friction is 0.22?

70. A skater allows himself to glide to rest and finds that he travels 60 yards in 12 seconds. If he weighs 180 lbs. What is the force of friction between the skates and the ice?

71. What is the magnitude of a force that, applied parallel to a plane inclined at 30° to the horizontal, will cause a block of 2 kg. mass to slide up the plane with an acceleration of 50 cm./sec.^{2}, if the coefficient of friction is 0.30?

72. A force of 50 lbs. is applied at an inclination of 30° to the horizontal to a block of 150 lbs. weight that lies on a horizontal surface and the coefficient of friction is 0.20. What acceleration is produced?

73. A uniform ladder that weighs 150 lbs leans without friction against a wall and its foot rests without slipping on a board. What is the greatest angle it can make with the ground and what is then the pressure on the ground if the coefficient of friction between board and ground is 0.50.

Machines.

74. A force of 1000 pounds is to be produced by means of a crowbar 4 ft. long by applying a force of 100 lbs. to the crowbar. Where must the fulcrum be?

75. What maximum weight could be raised by means of a jackscrew of $\frac{1}{8}$ inch pitch by a force of 100 lbs. applied at the end of a lever 1 ft. long?

76. What applied force would be necessary to raise a weight of 2 tons by a jackscrew of 0.04 inch pitch, the lever being 1 ft. long and the efficiency of the arrangement $\frac{3}{4}$?

77. A house-painter raises himself by a block and tackle with two pulleys in the upper block and a single pulley below. What force must be exerted on the rope if he weighs 200 lbs.?

78. The radii of a differential wheel and axle are 30 cm., 15 cm., and 12 cm. What applied force is required to raise a load of 1000 kg. and how much rope does the operator draw in while raising the load 4 meters?

Gravitation.

79. The balloon *Explorer II* rose to a height of about 13.7 miles on Nov. 11, 1935. How much less was the value of g there than at the earth's surface (radius of earth about 4000 miles)?

80. The mass of Mars is about $\frac{1}{10}$ that of the earth and the radius is about $\frac{1}{2}$ that of the earth. What is the value of g on Mars? If you could throw a baseball 100 yards on a level surface here, how far could you throw it on Mars?

81. The distance of the moon from the earth being taken as 240,000 miles and the mass of the earth as 80 times that of the moon, at what point between them would a meteorite be attracted equally by the two?

Elasticity.

82. If a sphere of steel of 10 cm. radius is subjected to a pressure of 100 kg. per cm², what is its decrease of radius?

83. Calculate Young's modulus for steel from the following observations: length of wire 200 cm.; diameter 1 mm.; stretching force 10 kg. wt.; change of length 1 mm.

84. What is the stretch produced in a copper wire 10 ft. long and of $\frac{1}{25}$ inch diameter by a weight of 10 lbs.?

85. One end of a wire 1 meter long and 1 mm. thick is clamped and the other end is twisted through 90°. What is the shear at the surface of the wire?

86. Two similar spheres, each of 40 g. mass, hang side by side from fine wires. One is drawn aside and released, so that it impinges on the other with a speed of 50 cm. per sec. What are their velocities after impact, if the coefficient of restitution is 0.8? Is momentum conserved? Energy?

Properties of Liquid.

87. A gold ring weighs 8.00 gms. in air and 7.45 gms. in water. How much silver is mixed with the gold?

88. To immerse a constant immersion hydrometer to the mark in water requires 175 gm. on the pan. What is the specific gravity of a liquid that requires 110 g., if the weight of the hydrometer is 200 gm.?

89. What weight can a block of ice of specific gravity 0.918 carry in salt water of specific gravity 1.036, if the volume of the ice is 10 ft.³?

90. To find the specific gravity of a body lighter than water it is weighed in air, giving 100 gm. wt., and then, with a second body added, in water, giving 70 gm. wt. The second body alone weighs 80 gm. in water. What is the specific gravity of the first body?

91. What weight of lead hung below a block of wood of 1 ft.³ volume will make it sink in fresh water, if the specific gravity of the wood is 0.75?

92. An anchor that weighs 200 lbs. in air is at a depth of 50 ft. in salt water of specific gravity 1.03. What amount of work is required to bring it to the surface?

93. There is a cavity in an irregular brass body that weighs 8.5 kg. in air and 7.0 kg. in water. What is the size of the cavity, if the density of the brass is 8.5?

94. Find the total thrust at the bottom of a barometric column 75 cm. high and 0.5 cm.² in cross-section.

95. Calculate the total thrust against a dam 40 yards long and 20 ft. wide, inclined at 30° to the vertical.

96. Two rectangular glass plates stand vertically in water with one pair of vertical edges in contact and the opposite pair slightly separated. What is the curve of the surface of the water between the plates?

97. Water rises to a height of 6 cm. in a glass capillary tube of 0.50 mm. diameter. Calculate the surface tension of water.

98. A jet of water from a small aperture in the side of a tank falls 25 cm. in travelling 100 cm. horizontally. How far is the aperture below the surface of the water in the tank?

99. Find the coefficient of viscosity of water from the following measurements: 235 c.c. of water flow in 10 minutes through a horizontal capillary

tube 25 cm. long and of 0.5 mm. radius when a constant pressure difference equal to that of 3 cm. of mercury is maintained between the ends.

Properties of Gases.

100. What percentage error is made in getting the density of cork (about 0.24 gm./cm.^{3}) from the weight and volume of a specimen, if no allowance is made for air buoyancy on the cork?

101. By what amount is the carrying power per cubic meter of a dirigible changed by substituting helium for hydrogen?

102. What is the difference in barometer reading between the bottom and top of a building 15 m. high? Why cannot the same method of calculation be used to find the height of a mountain?

103. To what height would the atmosphere extend if its density were everywhere the same 0.00129 gm./cm.^{3}, as at the surface, the height of the barometer being 76.00 cm. ("Height of homogeneous atmosphere").

104. Sometimes to measure the pressure in a very rarefied gas a sample is removed and compressed to a very small volume. In what proportion is the pressure increased by compressing a liter of the gas into a tube 10 cm. long and of 2 mm. internal diameter?

105. How high is water forced in a cylindrical diving bell 100 cm. high when it is lowered to the bottom of a lake 100 meters deep?

106. A barometer tube 80 cm. long contains a small amount of air above the mercury so that the reading is 74.00 cm. when a correct barometer reads 75.00 cm. What will this defective barometer read when the correct barometer reads 76.00 cm.?

107. A tube one meter long is one-third filled with water. It is then inverted so that its open end is under water. To what level will the water in the tube come?

108. A vessel of two liters capacity containing nitrogen at a pressure of two atmospheres is connected to a vessel of one liter capacity containing oxygen at a pressure of one atmosphere. What is the pressure when the gases have mixed?

109. What, according to the kinetic theory of gases, is the speed of a molecule in oxygen at standard pressure and density? Is this the average speed?

Wave Motion

Waves.

1. A body is vibrating with a period of 3.14 sec. and an amplitude of 2 cm. Find its displacement, velocity, and acceleration 15 seconds after it has passed through its equilibrium position.

2. A body hangs from a vertical spiral spring and vibrates vertically. If the force that the spring exerts on the body when the displacement is 2 cm. from the equilibrium position is twice that exerted in the equilibrium position, what is the period of vibration?

3. Calculate the amplitude of the resultant motion when two simple harmonic motions in the same line and of the same period are superimposed, if their amplitudes are 3 and 4 cms. and they differ in phase by $\pi/2$.

4. A particle has a motion that is the resultant of two simple harmonic motions at right angles. The components are of the same period and of the same

amplitude a, but one starts a quarter of a period before the other. Derive the equation of the path of the particle.

5. Two simple harmonic motions in the same line of the same amplitude and of periods 10 and 11 seconds start simultaneously from zero phase. After how many seconds will the resultant displacement be zero? When will this next occur?

6. Compound graphically two sine wave trains of wave-lengths as 2 to 1 and of amplitudes as 2 to 1, starting in the same phase.

7. Two sine waves of the same velocity, period, and amplitude are travelling on a string in opposite directions. At what points are the displacements always zero?

8. Calculate the velocity of a longitudinal wave in steel. Does the result you get apply to an unlimited mass of steel or to a steel wire?

9. The velocity of transverse waves on a wire of 1 gm. mass per meter is 140 cm. per sec. What is the tension in the wire?

10. What is the intensity at 1 meter distance of the spherical waves emitted from a small source in an isotropic medium if the power of the source is 1 watt?

11. The depth of the Cape Cod Canal at low water is 25 ft. At what speed does a long wave in it travel?

12. A passenger on a steamship travelling at 15 miles per hour observes that an ocean wave, (so long he can distinguish it clearly), seems to keep abreast of the vessel. What is its wave-length? Why can he not go on observing it?

13. A seismic sea wave, originating at Sanriku, Japan, reached Honolulu, distance 3500 miles, in 7 hr. 32 min. What does this give for the average ocean depth between the two places?

Heat

Temperature.

1. The highest known temperature in the shade was 135°F. (in Arabia) and the lowest known temperature was −88°F. (in Siberia). What are these temperatures on the Centigrade scale?

2. Find the temperature Fahrenheit of: the freezing-point of mercury, −38.89°C.; the melting-point of tin, 231.84°C.; the melting-point of platinum, 1764°C.; the surface of a class A star (*e.g.* Sirius), about 10,000°C.

3. Reaumur's thermometer is still widely used in Germany and some other European countries. It differs from the Centigrade thermometer in taking 80°R. as the boiling point of water. If a room temperature is 12°R. what is it on the Fahrenheit scale?

4. What is the boiling-point of water on the absolute Fahrenheit scale?

Expansion.

5. A metal rod is 100 cm. long at 20°C. and increases in length by 1 mm. when the temperature is raised to 100°C. What is its coefficient of expansion as defined in §273?

6. A copper telephone wire hangs between two posts 100 ft. apart. How much longer is it on a summer day when the temperature is 25°C. than on a winter day at −10°C.?

7. Steel rails are 40 ft. long at 60°F. What space must be left between them to allow for the expansion at 100°F.?

8. The Forth bridge in Scotland is of steel and is 5330 ft. long. What total expansion must be provided for (by roller bearings and rocking-posts) between 0°F. and 100°F.?
9. What is the proportional change of period of an uncompensated brass pendulum when the temperature changes by 10 degrees C.?
10. A vertical steel wire 1 mm thick carries a scale pan and a sensitive lever arrangement to indicate changes of length. When it is cooled 10 degrees C., what weight on the pan will stretch it to its original length? (See §171)
11. A circular disk of aluminum of 50 cm diameter at 10°C. is raised to 80°C. What is the increase in area of a face?
12. A mercury barometer stands at a height of 75 00 cm when the temperature is 20°C. What would its height be at the same atmospheric pressure, if the temperature of the mercury were 0°C.?
13. What is the relative coefficient of cubical expansion of mercury in Jena glass? How large a rise of temperature is required to make an apparent change of volume of 1% of the mercury in a glass thermometer?
14. The volume of a room is 90 cubic meters. What is the mass of the air in it at 20°C. and 75 cm barometric height if the density of air at 0°C. and barometric height of 76 cm is 0 001293?
15. When air is pumped into a tire the volume of the air is compressed to one fourth of its original volume and the temperature rises from 10°C. to 42°C. What is the pressure in the tire?
16. Picard (in 1932) rose to a height of 10 miles in a spherical aluminum shell 7.0 ft in diameter at 70°F. How much had its volume decreased when, at a height of 10 miles, the temperature was −70°F.?
17. In finding the depth of a lake by a sounding tube (problem 105, p. 683) it was found that the volume of the air at the bottom of the lake, where the temperature was 0°C. was 125 cm.^{3} How should this figure be corrected before calculating the depth, the temperature at the surface being 20°C.?

Calorimetry.

18. A calorimeter contains 200 gm. of water at 20°C When 200 gm. of water at 60°C. are poured in, the final temperature is 38°C. What is the thermal capacity of the calorimeter?
19. When 450 gm. of copper at 98 0°C. are dropped into 300 gm. of water at 18.0°C. contained in a copper calorimeter of 100 gm mass, the final temperature is 27 8°C. Find the specific heat of copper.
20. An iron block of 400 gm. mass is dropped from a furnace into 1250 gm. of water in a copper calorimeter of 110 gm. mass. If the temperature of the water rises from 20°C to 48°C. what was the temperature of the furnace?
21. If the heat of combustion of a fuel is 14,000 in B.T.U. per pound of the fuel, what is it in calories per gram?
22. It is found that the combustion of 1.200 gm. of coal produces enough heat to warm 580 gm. of water in a calorimeter of thermal capacity 280 from 17.7°C. to 28.9°C. What is the heat of combustion of this coal in B.T.U. per pound?
23. Gas that costs 70 cents per thousand cubic feet gives in burning 500 B.T.U. per cubic foot. What does it cost to have a bath in 5 cubic feet of water

heated from 50°F. to 95°F., if two-thirds of the heat of combustion goes into the water?

Mechanical Equivalent of Heat.

24. A lead bullet of 50 gm. mass is fired into a target with a velocity of 50 meters per sec. If the heat produced goes into the bullet, how much is its temperature raised?

25. Taking the height of Niagara Falls as 160 ft., calculate how much the temperature of the water should be raised by the drop, on the assumption that all the kinetic energy acquired by the water is transformed into heat in the water.

26. How much heat is developed when an automobile that weighs 4 tons and is travelling at 30 miles per hour is stopped by the brakes?

27. The power supplied to a drill in a quarry is 1.5 kilowatts (§61). How much heat is produced in a run of 2 minutes?

28. A flywheel of 2 ft. mean diameter and 200 lbs. mass is making 60 revolutions per minute. How much heat is produced when it is brought to rest by friction?

29. A body of 100 kg. mass slides without acceleration down a slope 50 meters long inclined at 30° to the horizontal. How much heat is produced by the friction?

Change of State.

30. To find the heat of fusion of ice a piece of ice of 3.00 gm. mass at 0°C. was dropped into a copper calorimeter of 100 gm. mass containing 91 gm. of water at 21.0°C. and the final temperature was 18.1°C. What did this give for the heat of fusion of ice?

31. When a kilogram of water at 100°C. mixes with and melts a kilogram of ice, what is the final temperature? What is the total change of volume when the ice has melted?

32. What is the ratio of the amount of steam at 150°C. and ice at −50°C. that will finally produce water at 50°C.?

33. To what fraction of its height (approximately) will a brass cylinder sink into a block of ice if its original temperature is 50°C.?

34. To what temperature would the air at 30°C. in a room $10 \times 10 \times 4$ meters be cooled by the evaporation of 1 kg. of liquid air?

35. How much work in ft.-lbs. would be required in rubbing two blocks of ice together to melt one ounce (§261)?

36. If the air in a room at 20°C. is saturated with water vapor, what is the mass of the water vapor in a cubic meter?

37. A room is $10 \times 8 \times 3$ meters and the temperature in it is 20°C. How much water vapor must be added to raise the relative humidity from 20% to 40%?

38. What is the relative humidity in a room at 30°C., if the dew-point is 20°C.?

39. What fraction of the heat of evaporation of water at 100°C. and under normal atmospheric pressure is expended in work against atmospheric pressure?

Heat Conduction.

40. A rectangular metal block of 20 cm. length and 15 cm.^{2} end area transmits heat in the direction of its length at the rate of 900 calories per minute when a temperature difference of 40°C. is maintained between its ends. Calculate the thermal conductivity of the metal.

41. If a thermal conductivity in calories per sec per cm per degree C. is to be reduced to B.T U.'s per sec. per inch per degree F. by what numerical factor should it be multiplied?

42. An iron steam-pipe is 2 meters long, 3 cm in internal diameter and 5 mm thick How much heat is lost from it per hour, if its outside surface is 15 degrees C. below its inside surface in temperature? To what mass of condensed steam does this correspond?

43. The average temperature gradient in the earth's crust is about one degree C rise per 100 ft. of descent (§325) Taking the average thermal conductivity as 0.004, calculate the thickness of a layer of ice that the upflow of heat would melt in a year

44. Make an approximate calculation of the amount of heat that flows in a minute through each sq. ft (take it as 1000 cm.^{2}) of ice 4 in. (10 cm) thick on a pond when the temperature of the air is -4°F. and that of the water 32°F and find by what fraction of an inch the ice thickens in that time Are the results likely to be too great or too small?

Radiation.

45. An incandescent lamp is raised in temperature from 3000°C to 3100°C. By what percentage is the emission of radiant energy from it increased?

46. The area of the surface of a steam radiator is 1 sq m. and the coefficient s in Stefan's law for the surface (§336) is 4×10^{-5} The walls of the room are at 68°F. At what rate does heat pass from the radiator to the walls, if the air temperature is practically constant?

47. A blackened copper ball of 200 cm ^{2} surface is heated to 300°C. and suspended in an exhausted enclosure the walls of which are also blackened and at 0°C. At what rate will it lose heat?

Thermodynamics.

48. The volume of a gas, originally under one atmosphere of pressure, increases from 1 liter to 2 liters. How much greater is the change of pressure if the expansion is adiabatic than it would be if the expansion was isothermal?

49. How much external work is done when a gas expands at a constant pressure of one atmosphere from 1 liter to 2 liters and what is the heat equivalent of this work in calories?

50. The density of oxygen at standard conditions is 0.001429, and its specific heat at constant pressure is 0.224 cal. per gm. per degree C. How much work is done when 1 gm. of oxygen is heated at constant pressure of 1 atmosphere from 0°C. to 1°C.? What value for the specific heat at constant volume follows from this?

51. When a gas has expanded adiabatically and thereby been cooled, its new pressure volume and temperature is connected with the former ones by the general gas law Find an equation for an adiabatic change in terms of volume and temperature.

52. Plot the Carnot cycle in a diagram with temperatures as ordinates and volumes as abscissae. (Note the changes in volume in Fig. 218.)

53. A gas under one atmosphere pressure expands adiabatically until the pressure is reduced to half an atmosphere. What is its volume? What final pressure is necessary to reduce it isothermally back to its original volume?

54. In problem 53 what was the ratio of the absolute temperature at the end of the expansion to that at the beginning?

55. What is the thermodynamic efficiency of an engine if steam enters the cylinders at 160°C. and is exhausted at 100°C.?

56. The engines of a ship run by turbines deliver 30,000 h.p. If the boiler efficiency is 80% and the heat of combustion of the coal is 15,000 B.T.U. per pound, how many tons of coal are required for a seven day voyage?

57. If coal sells at $10 a ton, what is the minimum cost of the energy delivered by a high-pressure turbine per kilowatt-hour (§61)?

Electricity and Magnetism

Electrostatic Fields.

1. Two very small, equally charged spheres, each of 0.2 gm. mass, are suspended from the same point by light silk fibres 100 cm. long. If, when they are in equilibrium, the centers of the spheres are 10 cm. apart, what is the charge in each?

2. Point charges of 40 and 60 e.s.u. are 50 cm. apart. Calculate the field intensity at the point midway between them and at a point on the line of the charges and 50 cm. beyond the larger.

3. What is the field strength at the vertex of an isosceles right-angled triangle, of 10 cm. side, when a charge of 100 e.s.u. is placed at each of the other two corners?

4. Two circular parallel plates, each of 4 cm. radius are 1.5 cm. apart and their potential difference is 2000 volts. What is the field-strength at a point midway between them and near the axis? If the slight distortion of the field near the edges is negligible, how many lines of force connect the plates? How much work is done by the field when one electron (4.8×10^{-10} e.s.u.) passes from one plate to the other?

Capacity.

5. What is the capacity of a condenser consisting of two parallel circular plates, each of 10 cm. diameter, 0.5 cm. apart in air?

6. The plates of a plate condenser are 10×20 cm., and they are separated by 2 mm. of glass. If the potential difference of the plates is 100 e.s.u., what is the charge on each plate?

7. How much work is done in charging a conducting sphere of 5 cm. radius to a potential of 100 volts?

8. If three capacities of 0.2, 0.2 and 0.05 microfarads are available, how can capacities of 0.45, 0.25 and 0.1 microfarads be made up?

9. A charged electrometer reads 1500 volts. When it is connected to a previously uncharged plate condenser of 150 e.s.u. capacity, the reading drops to 1000 volts. Disregarding the capacities of the connecting wires, calculate the capacity of the electrometer.

10. A charged plate condenser, with air as dielectric, is connected to an electrometer, which then reads 390 volts. When a second condenser, similar to the first but with an oil as dielectric, is joined in parallel with the first, the reading becomes 150 volts. What would this give for the dielectric constant of the oil of the capacity of the electrometer were neglected? What is the correct value, if the capacity if the electrometer is one-fourth of that of the first condenser?

11. A positive charge of 500 e.s.u. is placed at a distance of 50 cm. from an earthed sphere of 2 cm. radius. What charge appears on the sphere?

12. The radii of the plates of a spherical air condenser are 10 cm. and 10.2 cm. What is its charge when the potential difference is 100 volts?

13. Two spheres of 5 cm. radius have charges of 500 and 1000 e.s.u. Calculate their potentials. They are connected for a moment by a long fine insulated wire. Find their new potentials. What is the sum of the energies of the two charges before and after being connected?

14. What is the energy density in the dielectric of a mica plate condenser of 100 cm.^{2} plate area and 0.01 cm. plate distance, when the condenser is charged to a potential difference of 1.50 volts?

15. Find the capacity in microfarads of a meter length of a cable, regarding it as a cylindrical condenser of radii 1 cm. and 2.72 cm. with a dielectric of dielectric constant 4.

16. The capacity of a cylindrical condenser may, if the radii are nearly equal, be calculated by considering the condenser as equivalent to a plate condenser with slightly bent plates. Verify this by the following figures: $l = 100$ cm., $K = 1$, $r = 10.0$ cm., $r' = 10.1$ cm.

17. Calculate the surface density of the average charge of the earth in coulombs per cm.^{2} and in e.s.u. per cm.^{2}. To how many electrons per cm.^{2} does this correspond (§394)?

Magnetism.

18. What is the field-strength at a point on the axis of a bar magnet of magnetic moment 600 at a distance of 50 cm. from the center? What would the answer be if the pole strength were given as 100 and the distance of a pole from the center as 6?

19. Find the field-strength at a point on the perpendicular bisector of the axis of the magnet in the preceding problem, if the distance of the point from the axis is 50 cm.

20. A short, strongly magnetized iron bar of magnetic moment 600 is placed on a table with its axis in the direction of the magnetic meridian. A very short compass needle is moved in the line of the axis of the magnet and is found to reverse its direction at 20 cm. from the center of the magnet. What is the horizontal intensity of the earth's field?

21. A compass needle of magnetic moment 550 is at a point where the total intensity of the earth's field is 0.66 and the dip is 68°. What is the strength of the couple exerted on the needle?

22. A magnet on the bar of a magnetometer, 40.0 cm. from the center, produces a deflection of the needle of 10°. When the magnetometer is moved to a distance the magnet must be moved to 40.2 cm. from the center to produce the same deflection. What is the ratio of the strengths of the two fields?

23. A compass needle makes 8 complete vibrations in a minute on a field out of doors. When it is carried into a laboratory the number of the vibrations becomes 7.5 per minute. In what proportion is the field weaker in the laboratory?

24. A slender magnet of 5 cm. length and 24 gm. mass makes 7.5 vibrations per minute when vibrating as a compass needle. When it is used as the deflecting magnet in a magnetometer it produces a deflection of 10° at a distance of 25 cm. What is the horizontal component of the earth's field?

25. A small magnet when mounted as a compass needle makes 5 vibrations per minute, and when mounted as a dipping needle it makes 10 vibrations per minute. Calculate the dip.

26. The elements of the earth's field at a place are: horizontal intensity 0.168, declination 12.2° W., inclination 66.3°. What is the strength of the field, and what is its horizontal component in the geographic meridian?

Heat Produced by Currents.

27. What current in a coil of 0.300 ohms resistance will develop enough heat in 10 minutes to heat 60 gm. of water in a copper calorimeter of 100 gm. mass 10.0 degrees C.?

28. How many calories does an electric heater of 20 ohms resistance on a 110 volt line develop in a minute?

29. With electric energy selling at 5 cents per kilowatt-hour, what is the cost of a million calories? What do a million calories cost if produced from coal at $12 a ton, the heat of combustion being 8000 calories per gm.?

Magnetic Fields of Currents.

30. What is the field strength at the center of a coil of 38 cm. diameter carrying a current of 4 amperes?

31. What is the strength of the couple exerted on a short horizontal bar magnet of magnetic moment 500, when it is 30 cm. from the center of a long vertical straight wire that carries a current of 2 amperes, its axis being perpendicular to a line from its center to the wire?

32. It is desired to produce a magnetic field strength of 40 e.m.u. near the center of a solenoid of 1 meter length by means of a current of 5 amperes. With how many turns must the solenoid be wound?

33. A toroidal solenoid 40 cm. in mean diameter and wound with 920 turns of wire carries a current of 5 amperes. How many magnetic lines of force pass through 2 sq. cm. within the solenoid in a plane through the axis?

34. A magnet of 4 cm. length and pole strength 20 e.m.u. kept horizontal is carried around a long vertical wire that carries a current of 5 amperes, so that its center moves in a circle of 20 cm. radius and the magnet always points at the wire. Calculate the work done on each pole separately. What difference would it make if the magnet always pointed magnetically N and S?

35. At a station where the horizontal component of the earth's field is 0.168 its effect on a small magnetic needle is to be neutralized by the current in a circle of wire of 15 cm. radius. How must the circle be placed and what current will be needed?

36. A current of 15 amperes flows in a straight solenoid 60 cm. long and 3 cm. in diameter. How many lines of force emerge from an end face and how many leak out near an end?

Galvanometers.

37. The deflection of a tangent galvanometer of a single turn of wire of 18.5 cm. radius is 32°, and H is 0.158 e.m.u. What is the current?

38. The scale of a tangent galvanometer is sometimes graduated to read amperes directly. If such a galvanometer has 8 turns of wire, what should its radius be so that the scale shall read amperes directly, if $H = 0.178$?

39. A coil of 20 turns of 35 cm. diameter is in a vertical plane at 30° with the magnetic meridian where H is 0.172. A small compass needle at the center comes to rest in the plane of the coil. What is the current in amperes in the coil?

Electric Resistance.

40. What is the resistance of a copper wire 0.65 mm. in diameter and 100 meters long?

41. The resistance between two points 15.0 cm. apart on a circular iron bar of 1.00 cm. diameter is 200×10^{-6} ohms. What is the specific resistance of the iron?

42. The resistance of a wire at 20°C. is 2.70 ohms and at 100°C. it is 3.50 ohms. What is the temperature coefficient of the material?

43. At what temperature is the resistance of a platinum wire twice as great as at 20°C.?

44. Coils of 2, 3, and 6 ohms resistance are in parallel. What is the equivalent resistance of the arrangement?

45. How will a current of 1.8 amperes be divided between three resistances of 1, 3, and 6 ohms in parallel?

46. A galvanometer of 100 ohms resistance is connected in parallel with a wire of 1000 ohms resistance. A circuit is connected to one terminal of the galvanometer and to the middle of the wire. What is the equivalent resistance between the terminals?

47. What is the cost of lighting a room with two 60-watt lamps and one 25-watt lamp for 6 hours, if the electric energy costs 5 cents per kilowatt-hour?

48. An electric lamp takes 60 watts on a 110-volt circuit. What is the resistance of the lamp and the current through it?

49. How many 60-watt lamps can be used for the same length of time and at the same cost as a flatiron that takes 5.35 amperes at 112 volts?

50. A string of 8 similar small lamps in series, used to light a Christmas tree, takes as much power as a single 50-watt lamp on a 110-volt circuit. What is the resistance of each small lamp? If a boy, tinkering with a broken string, managed to shunt two of the small lamps in parallel onto the 110-volt line, what current would pass through the house fuse?

51. When the terminals of a cell of 1.1 volt e.m.f. are connected by a 5-ohm wire, the potential difference of the terminals falls to 0.6 volt. What is the resistance of the cell, and what current flows in the wire?

52. The current in a 1-ohm wire connected to the terminals of a cell is 0.1 amperes. When the 1-ohm wire is replaced by a 12-ohm wire, the current drops to 0.05 ampere. What is the resistance of the cell? Its e.m.f.?

53. What shunt resistance should be used with a galvanometer of 118 ohms in which a current of 2.45×10^{-7} ampere produces a 1 cm. scale deflection to make it a direct-reading milliammeter?

54. What resistance is needed and how should it be used to make a direct-reading voltmeter out of the galvanometer of problem 53?

55. A high resistance voltmeter is connected across the terminals of a small generator serving a summer camp. With the generator running at the proper speed but without any lamps being turned on, the voltmeter reads 110 volts. When three lamps, of 210 ohms resistance each, are turned on

in parallel the voltmeter reads 105 volts. What is the current in each lamp? What is the equivalent resistance of the generator coils?

56. A battery consists of 16 cells, each of 2.00 volts e.m.f. and 0.1 ohm resistance, arranged in two parallel lines of 8 cells each. The battery furnishes current to a resistance of 2 ohms in series with 2, 4, 4 ohms in parallel. If the total resistance of the lead wires is 0.6 ohm, what current flows in a 4-ohm coil?

57. A steady current is maintained in a manganin measuring wire of 0.30 mm. diameter. A Daniell cell (e.m.f. 1.105 volts) with a galvanometer in series is connected at two points on the wire 56.7 cm. apart and the galvanometer shows no deflection. What is the current in the wire?

58. A current passes through an ammeter and through an accurately measured resistance of 0.500 ohms, in parallel with which are two large resistances of 3000 and 8000 ohms. A Daniell cell (e.m.f. 1.105 volts), with a galvanometer in series, is shunted across the 8000 ohms and the galvanometer is not deflected. If the ammeter reads 3.00 amperes, is it correctly calibrated?

59. The charge on a condenser of 1 microfarad capacity decreases in 10 minutes to a fraction 1/2.72 of its original value. Through what resistance is it discharging?

60. The e.m.f.'s of two cells are 2 volts and 1 volt respectively, and their respective resistances are 2 ohms and 1 ohm. They are joined in parallel by wires, of negligible resistance, and used to produce a current in a wire of 1 ohm resistance. Apply Kirchhoff's laws to find the current in the wire and the current through each cell.

Magnetic Induction.

61. The intensity of magnetization of an iron specimen in a magnetizing field of 10 oersted is 100. Calculate the magnetic induction, the permeability, and the susceptibility of the specimen.

62. An iron ring of 0.5 cm.^{2} cross-section and 10 cm. diameter is wound with 300 turns carrying a current of 1 ampere. What is the induction in the iron if the permeability at the existing value of H is 1000?

63. A rod of permalloy of 4 mm.^{2} cross-section is pointed in the direction of the earth's field where the strength of the field is 0.65 oersted. If the permeability is 75,000, what is the induction, the intensity of magnetization, and the pole strength?

Thermoelectricity and Electrolysis.

64. Determine from the thermoelectric diagram (Fig. 380) the thermal e.m.f. of an iron-copper couple, one junction being at 0°C. and the other at the neutral temperature.

65. The thermo e.m.f. of a copper-constantan couple is, over a wide range, 40×10^{-6} volts per degree C. What deflection will be produced by putting such a couple, with 0.01° degree difference of temperature at the junctions, in series with a galvanometer of 100 ohms resistance, which gives 1 cm. scale deflection for a current of 2.5×10^{-7} amperes?

66. A steady current deposits 2.48 gm. of silver in 15 minutes. What is the current?

67. Owing to a flood a stray current of 5 amperes gets onto a lead pipe. How much lead is eaten away in a week where the current goes from the lead to moist earth?

68. How much lead changes to lead sulphate per ampere-hour in a lead storage cell?

69. The plane of a square coil of 30 cm. side and carrying a current of 2 amperes makes an angle of 30° with a magnetic field of 5 e.m.u. What is the magnitude of the couple exerted on the square?

Electromagnetic Induction.

70. A rectangular coil of a d.c. motor armature is 25 cm. long and 15 cm. wide and it carries a current of 10 amperes. The field strength is 3000 oersteds. What is the maximum torque exerted on the coil?

71. A copper disk of 6 cm. diameter is at right angles to a magnetic field of 100 oersteds and is rotated about its axis 60 times per sec. What e.m.f. is induced between its center and a point on its circumference (Faraday's disk dynamo)?

72. The coil of an earth inductor consists of 50 turns of wire of 15 cm. diameter and its resistance is 20 ohms. What quantity of electricity in coulombs, is induced in it when it is turned from a position in a vertical plane perpendicular to the magnetic meridian, where $H = 0.18$, to a position parallel to the magnetic meridian?

73. A circular coil of 40 turns of 20 cm. radius rotates 60 times per second about an axis through a diameter perpendicular to a magnetic field of 12 oersted strength. What is the maximum of the induced e.m.f.? In what position of the coil is the instantaneous e.m.f. half of the maximum?

74. What is the coefficient of mutual induction of two solenoids, 1 meter long and each of 1000 turns, one being wound on the other and the mean radius being 2 cm.?

75. What is the impedance of a coil of 100 ohms resistance and 100 millihenries inductance when a sine current of frequency 60 flows in it? If the maximum e.m.f. is 110 volts what is the current?

76. A solenoid of 100 cm. length and 6 cm. diameter, wound with 1200 turns, carries a current of 2 amperes. Near its center a secondary of 200 turns and 2 ohms resistance is wound and is connected to a ballistic galvanometer of 118 ohms resistance. What discharge passes through the galvanometer when the current in the primary is reversed?

77. In what time will the current in a coil of 100 millihenries inductance and 50 ohms resistance rise to 99% of the value given by Ohm's law when a constant e.m.f. is applied to it?

78. The "half-life period" of radium A is 3 minutes. What length of time does it take for 90% of the atoms of radium A in a specimen to disintegrate?

Radioactivity, etc.

79. How many electron-volts are equivalent to one erg?

80. What is the velocity of an electron if its energy is 1 electron-volt?

81. Verify the statement in the text (§548) that the energy, 1.2×10^{-5} ergs, of an α-particle is equal to 7.66 million electron-volts.

82. What is the maximum speed with which an electron can escape from a metal surface for which the threshold value of wave-length is 390×10^{-7} cm. and the work function is very small?

83. What is the "mass" of a photon of sodium light? How does it compare with the mass of an electron?

84. If the conversion of mass into energy (§552) could be carried out completely, how much energy could be derived from a gram of any substance? Convert the result into calories and find what amount of coal would yield the same amount of heat, if the heat of combustion is 7800 calories per gram.

85. Calculate the energy of an α-particle moving with a velocity of one fifteenth of the velocity of light; of a β-particle with a velocity of one third that of light (without the relativity correction, which would be small); of a γ-ray of wave-length 10^{-9} cm.

86. What strength of magnetic field is required to compensate an electric field of 1000 volts per cm., acting on cathode rays the velocity of which is one-tenth of that of light (§537)?

Sound

1. In what proportion does the velocity of sound decrease in a thunderstorm when the temperature falls from 32°C. to 16°C.?

Sound.

2. Compute the velocity of sound in air at 0°C. and a barometer height of 76 cm., if the ratio of the specific heats of air is 1.405.

3. The velocity of sound in helium is found to be 936 meters per second. What is the ratio of the specific heats of helium?

4. The pitch of a locomotive whistle dropped from C to B as the locomotive went past a stationary observer. What was the speed of the locomotive?

5. The first overtone of a stopped organ pipe of 100 cm. length is in resonance with a tuning-fork of frequency 258. What is the velocity of sound in air?

6. A piano tuner counts 2 beats per second between his standard A tuning fork of frequency 440 and the corresponding piano wire. In what proportion is the tension of the wire to be changed?

7. On an ordinary piano there are seven octaves and two notes. The frequency of the A at the middle of the piano ("middle A") is 440. What is the frequency of the lowest note, A_4 (which is four octaves below middle A) and that of the highest note, C'''', (which is four octaves above "middle C").

8. An open organ pipe is 150 cm. long. What are the frequencies of its fundamental and its first overtone at 0°C.? At 18°C.?

9. The average distance between two dust heaps in a Kundt tube is 3.40 cm. when the tube is filled with air and 3.10 when the tube is filled with argon. What is the velocity of sound in argon? Does this check with the value for helium in problem 3?

10. If in problem 9 the length of the distance between the clamped points on the rod is 47.6 cm., what is the velocity of sound in the glass? Calculate Young's modulus of elasticity for the glass.

11. What is the depth of water under a ship 450 ft. long, if the sound of the detector at the bow, seems to come from a direction 40° below the horizontal?

12. What percentage change of power of a source of sound of constant pitch, for example a fog-horn, is necessary to increase the loudness (sensation level) for that pitch by 1 decibel? By how many decibels would doubling the power increase the loudness?

13. A lecture room is 9 meters by 10 meters and 4 meters high. The floor is of wood, coefficient of absorption .06. The coefficient of absorption of walls and ceiling, chiefly plaster, may be taken on the average as .03. What is the reverberation time when the room is empty? When there is an audience of 50? How many square meters of absorbing felt, coefficient of absorption .78, should be placed on the walls to make the reverberation time, with an audience, 1.2 sec.?

Light

Reflection.

1. The scale for reading the rotation of a magnetometer mirror is 120 cm. from the magnetometer mirror. The scale reading (as observed through a small telescope) changes by 20 cm. Through what angle has the magnet turned?

2. When the sun is 60° high, sunlight is brought horizontally into a room by reflection from a mirror M followed by reflection from a mirror M'. If M' makes an angle of 45° with the horizontal what is the tilt of M?

3. An "optical lever" is often used for measuring small changes of length of a rod due for example to heating. As an illustration suppose that on a small brass plate supported on three sharp-pointed legs at the corners of an equilateral triangle of 2 cm. side there is a small mirror with its plane perpendicular to the plate and parallel to the line joining two legs. If the third leg is lifted 0.1 mm., what change is there in the reading on a scale 2 meters away, as observed by a telescope directed at the mirror?

4. A small incandescent lamp is between two parallel mirrors 40 cm. apart and at a distance of 15 cm. from one mirror. How far behind this mirror are the first two images of the lamp?

5. When you are walking toward a plane mirror at what speed does your image seem to come to meet you?

6. When you are driving a car and see in the mirror another car coming up behind you, is the speed at which he seems to approach you his real speed of approach?

7. Plot roughly a curve showing the general relation between object-distance u and image-distance v, (a) for a concave mirror, (b) for a convex mirror.

8. A nail 2 in. long is held at a distance of (a) 60 in., (b) 15 inches from a concave mirror of 50 in. radius. Find the position, nature and size of the images and construct them graphically.

9. An object of 1 cm. length is placed at a distance of 30 cm. from a convex mirror of 20 cm. radius. Find the position, nature and size of the image and construct it graphically.

10. If a writer of a text book found a picture that he would like to put in his book, but on a smaller scale, how might he find, by means of a convex mirror, what it would look like if reduced to one half in linear dimensions? Could he use a concave mirror equally well?

11. The radius of curvature of each face of a double convex lens is 30 cm. It is held so that images of a distant window are seen in the two faces. How can the image from the rear face be distinguished from that from the front face? Where does the latter image appear to be?

12. The full moon subtends an angle of 31′ at the earth. How large a photographic plate in the focal plane of a concave reflector of 200 ft. radius of curvature is required to obtain a complete picture of the moon?

Refraction.

13. A rod stands obliquely in water at an angle of 30° with the normal to the surface. A second rod is held above the surface so that it is in line with the immersed part of the first, and it makes an angle of 42.5° with the normal to the surface. What is the index of refraction of water?
14. A glass plate 2 mm. thick ($n = 1.58$) is covered by a layer of water 2 mm. deep. What is the equivalent air path? What would it be if the water were replaced by oil ($n = 1.58$)?
15. After a microscope has been focussed on an object, a cover glass 1.50 mm. thick of crown glass ($n = 1.53$) is placed over the object. How much must the microscope be raised to remain focussed on the object?
16. A ray of light falls at an angle of incidence of 30° on a glass plate 2.5 mm. thick ($n = 1.60$). What is the distance from the point of incidence on the first surface to the point in which the ray reflected from the second surface emerges from the first surface? What is the distance between the two rays after reflection and refraction?

Lenses.

17. The radius of curvature of the curved surface of a plane-convex lens is 20 cm. and the index of refraction of the glass is 1.56. What is the focal length of the lens?
18. Light falling normally from above on a watch glass of 10 cm. radius of curvature filled with an oil is focussed at 25 cm. below. What is the index of refraction of the oil?
19. What is the focal length of a concavo-convex lens of glass of index of refraction 1.60, if the radii of curvature of the faces are 25 cm. and 15 cm.?
20. The focal length of a convexo-concave lens is found by experiment to be 25 cm. and the radii of curvature of the faces are found to be 20 cm. and 40 cm. What is the index of refraction of the glass?
21. Find the position, nature and size of the image of an object 1.5 cm. high, placed 100 cm. from a converging lens of 40 cm. focal length. Construct the image graphically.
22. An object 2 cm. high is placed 20 cm. from a converging lens of 25 cm. focal length. Find the position, nature and size of the image and construct it graphically.
23. Plot roughly a diagram showing the relation between u and v for a converging lens, taking object-distances u as abscissas and image-distances v as ordinates.
24. Plot roughly a diagram showing the relation of u and v for a diverging lens.
25. The planet Venus at its closest approach to the earth subtends a visual angle of about 1′. How large is the image of Venus formed in the focal plane of the Lick observatory refractor the focal length of which is 18 meters?
26. Two converging lenses, each of 20 cm. focal length, are 60 cm. apart. Where is the final image of an object 40 cm. from one and 100 cm. from the other?
27. A source of light 2 cm. high is placed 30 cm. from a converging lens of 20 cm. focal length. A second converging lens, also of 20 cm. focal length

is 40 cm. beyond the first. Calculate the position and size of the final image.

28. Where and how large is the final image, if the second lens in the preceding problem is replaced by a diverging lens of 15 cm. focal length?

29. The radius of curvature of the curved face of a hemispherical lens ($n = 1.55$) is 1.5 cm. The plane face is in contact with oil of the same index of refraction. Where must an object be placed in the oil so that a distinct image of it can be seen above the lens? Where will the image be?

30. What is the effective focal length of a converging lens ($n = 1.48$) if it is surrounded by carbon bisulphide ($n = 1.67$), if its focal length in air is 30 cm.?

31. A horizontal scale is seen in a telescope by reflection from a vertical mirror, the objective of the telescope being just above the scale. The focal length of objective and eyepiece are 8 cm. and 2.5 cm. For what distance between them will an image of the scale be seen at a distance of 25 cm. from the eyepiece?

32. A double convex lens of crown glass ($n_C = 1.55$, $n_F = 1.56$) is to be achromatized by fitting to it a plano-concave lens of flint glass ($n_c = 1.66$, $n_F = 1.68$). If the focal length of the doublet is to be 20 cm. what radii of curvature are required?

Photometry.

33. A lamp of unknown candlepower is at one end of a photometer bench 300 cm. long, and a standard 16 candlepower lamp is at the other end. If the screen is adjusted to equal illumination at a distance of 140 cm. from the second lamp, what is the candlepower of the first?

34. At what point between two lamps, the candlepowers of which are as 1 to 4, would the photometer screen be equally illuminated by the two lamps, the length of the bench being 300 cm.?

35. If 1 lumen is taken as equivalent to 0.0015 watt, what is the intensity corresponding to 1 lux in calories per cm.^{2} per sec.?

36. A lamp hangs 3 ft. above a desk. If a book lying on the desk and originally directly below the lamp is displaced 1 ft. sidewise, in what proportion is the illumination of the page decreased?

37. A plane mirror and a screen are placed parallel and 100 cm. apart. A small source of light is 80 cm. from the screen and 20 cm. from the mirror. In what proportion is the intensity of illumination on the screen at the point nearest to the source increased by the presence of the mirror?

38. What would be the answer to the preceding problem if the plane mirror were replaced by a concave mirror of 20 cm. focal length?

Dispersion.

39. The angle of minimum deviation of a 60° prism for sodium light is 38° 30′. What is the index of refraction of the prism for that color?

40. What is the angle of a prism of heavy flint glass that will achromatize a prism of crown glass with an angle of 8° for the C and F lines?

41. What is the dispersion of a 30° prism between the red ($n = 1.63$) and the blue ($n = 1.61$) when white light falls normally on one face?

42. A glass plate ($n = \frac{3}{2}$) is covered with water ($n = \frac{4}{3}$). What is the critical angle for light coming from the glass?

43. Light enters normally through the base of an isosceles prism of crown glass and is totally reflected at the sides. What is the minimum angle required at the apex of the prism for this to happen?

Interference.

44. A wedge of air is formed between two glass plates 10 cm. long in contact at one side and separated at the other by a wire 1 mm. thick. How far apart are the bright lines when red light falls normally on one plate?

45. The radius of curvature of the curved face of a plano-convex lens is 6 meters. When this face is pressed firmly against a plate glass surface and light falls normally on the plane face of the lens, the radius of the 64th interference ring is 1.50 cm. What is the wave-length of the light?

46. What would be the radius of the same ring (preceding problem) if the air were replaced by water?

47. The slits in a two-slit experiment are 1 mm. apart. How long is the first-order colored band, from violet to red, on a screen 3 meters away?

48. Light from a narrow slit at a distance of 25 cm. passes through a Fresnel biprism ($n = 1.60$), the angles of which are 12′. What is the angle between the emerging rays? What is the equivalent slit distance 25 cm. from the prism? What is the distance between two interference maxima on a screen 1 meter beyond the prism?

49. When a glass film 0.010 mm. thick is placed in the path of one of the interfering beams in an interferometer and causes a retardation of 9 wave-lengths of sodium light, what is the index of refraction of the glass for this light?

Diffraction.

50. The angular deviation of the third order spectrum produced by a transmission grating with 110 lines per mm. is 11° 11′. What is the wave-length of the light?

51. Parallel light falls normally on a plane reflection grating with 1000 lines per mm. In what direction is the first-order spectrum of sodium light reflected from the grating? What is the answer if the angle of incidence is 30°?

52. Two screens S and S' are 5 meters apart, and in S there is a circular aperture 2.42 mm. in diameter. Sodium light falls normally on S. What intensity does it produce at the point P on S' that is on the axis of the aperture? Where should an opaque disk of 1.71 mm. diameter be placed in the path of the light to produce maximum illumination at P?

53. Sodium light passes through a slit 0.2 mm. wide and falls on a photographic plate 300 cm. away the distance between the center of the most prominent black line and the next is 8.85 mm. What is the wave-length of the light? What distance would have to be measured on the plate, if the experiment were to succeed with x-rays?

54. What is the resolving power for yellow light of the large refracting telescope in the Yerkes observatory if the diameter of the lens is 102 cm.

Polarization.

55. If light from a sodium flame falls on glass of index of refraction 1.57, at what angle of incidence is there a maximum of polarization in the reflected light?

56. Light passes in succession through two Nicol's prisms. With what fraction of its intensity at incidence on the second prism does the light emerge,

if the angle between the principal sections of the prisms is 0°? 30°? 45°? 60°? 90°?

57. What is the thickness of a quartz plate if it is a quarter wave plate (§§751, 743.)

58. A cane sugar solution in a tube 20 cm. long produces a rotation of the plane of polarization of 7°. What is the concentration of the solution?

59. What thickness of quartz will compensate the rotation produced by a 10 % cane sugar solution in a tube 20 cm. long?

Spectrum.

60. An incandescent black body begins to glow at about 700°C. At what wave-length does the maximum intensity occur?

61. How accurately must the shift of the hydrogen line C be measured to detect a relative motion of a star toward the earth of 10 km./sec.?

INDEX

Numbers Refer to Pages.